DOCUMENT RESUME

ED 192 466 EA 013 086

AUTHOR	Boruch, Robert F.; Cordray, David S.
TITLE	An Appraisal of Educational Program Evaluations: Federal, State, and Local Agencies.
INSTITUTION	Northwestern Univ., Evanston, Ill.
SPONS AGENCY	Department of Education, Washington, D.C.
PUB DATE	30 Jun 80
CONTRACT	300-79-0467
NOTE	419p.
EDRS PRICE	MF01/PC17 Plus Postage.
DESCRIPTORS	Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; Federal Aid; *Federal Programs; *Federal Regulation; Government (Administrative Body); Government Role; *Program Evaluation; School Districts; State Departments of Education

ABSTRACT

This report concerns evaluation of federally supported educational programs at the national, state, and local levels. It was undertaken in response to Section 1526 of the Education Amendments of 1978, which requires that the Commissioner of Education conduct a comprehensive study of evaluation practices and procedures. Two broad sources of information were used: contemporary research and development by other researchers, and direct investigations by the project staff. Introductory material is presented in the first chapter. Chapter 2 considers the rationale, evidence, and opinion bearing on why evaluations are done; the confusion and argument engendered by general demands for evaluation; and the audiences to whom evaluations are addressed. Chapter 3 addresses the question of how evaluations are executed. Chapter 4 covers the organization of evaluations and the capabilities of evaluators, and chapter 5 considers quality of evaluations. The way evaluation results are used is considered in chapter 6, and case studies on the use of evaluative information are included. Chapter 7 covers recommendations. An extensive bibliography concludes the report. Legislative and management background, and research strategies are contained in the appendixes. (Author/MLF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED192466

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

An Appraisal of Educational Program Evaluations: Federal, State, and Local Agencies

Northwestern University

Evanston, Illinois

Principal Investigators: Robert F. Boruch
David S. Cordray

Submitted to: Department of Education

Date: June 30, 1980
Contract Number: 300-79-0467
Contract Title: A Comprehensive Study of Evaluation Practices and Procedures in Federally Funded Elementary and Secondary Education Programs

EA 013 086

OCT 14 1980

EXECUTIVE SUMMARY FOR THE CONGRESS

→ We conclude that the statutory provisions for evaluation are often unclear and that formal mechanisms for clarifying evaluation goals are necessary. We recommend that evaluation statutes identify the specific questions which need to be addressed and specific audiences for results, and that evaluability assessments be undertaken where specification is not possible. We recommend regular joint planning by Department and Congressional staff to clarify Congressional information needs and to decide when evaluation is warranted.

→ Evaluation activities at LEA and SEA levels vary considerably. The capabilities required of evaluators depend heavily on these local evaluation tasks. We recommend then that new demands for evaluation aimed at LEAs and SEAs be preceded by a "capabilities assessment" to determine whether time, money, and personnel exist to adequately execute the tasks demanded. We recommend expansion of training and technical assistance when evaluation demands are high.

→ Good evaluation designs help to prevent expensive mistakes. But they are not used often, partly because innovations are planned independent of evaluation. We recommend that pilot tests be undertaken before new programs or program variations are adopted and that the introduction of new programs be staged so that good designs can be exploited. Further, we recommend that higher quality evaluation designs, especially randomized experiments, be authorized explicitly in law for testing new programs, new variations on existing programs, and new program components.

→ There is a need for independent, balanced critique of major program evaluations before the evaluation results are used in policy. We recommend routine critique of major national evaluations required by law, periodic critique of samples of evaluations generated at the local, state, and federal levels, and statutory provisions for making statistical data available for reanalysis.

→ We conclude that the arguments over whether evaluations are used are frequently uninformed, that evaluations have been used, but that their use can be improved. We recommend regular discussion between evaluation staff of the Department and Congressional staff to facilitate use, and routine identification of useful evaluations in Congressional reports and in Department policy statements.

→ Recently developed standards and guidelines on evaluation are inappropriate for incorporation into law. They can be used to guide screening of evaluation proposals submitted by SEAs and LEAs in competitive grant programs. And they are a reasonable basis for reaching nontechnical agreements at the federal level, about what can reasonably be expected in an evaluation.

EXECUTIVE SUMMARY

FOR THE DEPARTMENT OF EDUCATION

→ We conclude that although restrictions on program staffers initiating conversations with Congressional staff may be warranted, restrictions on evaluation unit staff are not. Restrictions do not foster better understanding of Congressional needs for information or better planning, and ultimately they probably degrade utility of evaluations. We recommend that no such restrictions be placed on the unit.

→ We conclude that the absence of regular discussion with Congressional staff to plan evaluations degrades timeliness, relevance, and credibility of evaluations. We recommend that the Department create formal policy on this, and vigorously continue recent efforts to establish a regular dialogue.

→ High quality evaluation designs for estimating the effects of programs on children are not used often. We recommend that the Department authorize explicitly the use of randomized field experiments to plan and evaluate new programs, new variations on existing programs, and program components.

→ We conclude that there is a need for independent critique of evaluation reports produced at SEA, LEA, and federal levels. We recommend periodic sampling of LEA and SEA reports to provide balanced independent critique. And we recommend formal provision in contracts and policy for independent critique of major federal efforts.

→ We conclude that access to and identification of evaluation reports is often inadequate. We endorse recent action to limit the clearance period for reports to ten days and recommend adherence to the new limit. We also recommend adherence to a more conscientious system of specifying documents, identifying core recipients of reports, and acknowledging authors.

→ We conclude that existing methods of tracking use of evaluation results in the Department are poor. We believe that reporting use in the Annual Evaluation Report is desirable, but it can be improved by making reference more specific. We recommend that a formal system be created to track use of results by program managers and by the Congress. The earlier recommendations on dialogue between Congressional staff and evaluation unit staff should help to assure that more useful evaluations are mounted.

→ New programs are often badly implemented and we know considerably less than we should about the implementation process. We recommend formal intensive measurement of the degree to which activities match plans where measurement techniques are available. We recommend adjoining research on methods of measuring implementation to the implementation process because measurement methods are often not available.

DIGEST: RECOMMENDATIONS AND RATIONALE

Recommendations for the Congress are discussed first. Recommendations to the Department follow.

Planning and Executing Evaluations

We recommend that the Congress direct the relevant staff of Congressional committees, GAO, and CBO to meet with evaluation staff of the Department regularly to:

- reach agreement about when particular evaluations are warranted, and the senses in which each evaluation required by law is possible.
- clarify Congressional information needs, quality of evidence required, and planning cycle for each major evaluation required by law.
- identify specific committees and groups as audiences for evaluation results.
- identify the changes in program or understanding which could occur on the basis of alternative findings.

This recommendation hinges partly on the fact that a statutory demand for "evaluation" is ambiguous. The word can imply any activity from jouralistic reporting to full-blown field experiments dedicated to estimating the effects of innovation on children. The involvement of multiple interest groups is often necessary, but complicates matters. At worst, general demands to evaluate obscure the fact that feasibility of evaluation varies enormously and that elaborate evaluation may be unnecessary. Periodic efforts have been made by members of the Congressional staff to assure that production of evaluations coincides with authorization cycles and that Congressional needs are understood. The process is less regular and less orderly than it ought to be.

Statutory Provisions for Evaluation

We recommend that in constructing statutory provisions for evaluation that the Congress:

- specify exactly which questions ought to be addressed and the audiences to whom results should be addressed.
- provide for formal assessment of the evaluability of the relevant program where specification is not possible.
- provide for statistically valid field testing of proposed evaluation requirements where specification is not possible and in-house assessment insufficient.

Though statutes are explicit about routine reporting requirements, references to evaluation often are not specific. The simple requirement to evaluate whether the program meets objectives of the statute is common and vague. Hearings are often not informative. Defining evaluation requirements in terms of the questions which should be addressed is sensible so long as the questions themselves make sense, answering them is feasible, and the answers are likely to be useful. The specification of audiences, especially particular committees or Congressional support agencies, should enhance usefulness. We recognize that explicitness is often not feasible or desirable. Consequently, we suggest formal investigation of evaluability to clarify questions, audiences, and the ways in which results can be used, within a year after enactment of a demand for evaluation.

Evaluator Capabilities

We recommend that

- capabilities be assessed before new statutory evaluation requirements are directed at LEAs and SEAs to determine where resources are adequate to meet the demand
- expansion of training or technical assistance when the demands are notable and capabilities low
- explore the feasibility and desirability of direct contracts programs to capitalize on LEA and SEA capabilities.

The first recommendation stems from conclusions that no real standard for assigning the title "evaluator" exists. Skills required of the evaluator depend heavily on nature of the evaluation demand and on LEA and SEA interest in evaluation. The second recommendation is based on the finding that most LEAs and SEAs need assistance when the demand is high and want it. A small minority of LEAs have strong evaluation units. But these are a major resource and we believe that direct grant opportunities should be expanded to capitalize on them.

Use of and Authority for Better Evaluation Designs

We recommend that the Congress:

- routinely consider pilot testing every new program, variations on existing programs, and program components before they are adopted at the national level, using high quality evaluation designs.
- authorize the Secretary explicitly in each evaluation statute to use high quality designs, especially randomized field experiments, for planning and evaluating new program components, program variations, and new programs.

The rationale for the first recommendation is that higher quality evaluations are more feasible before the program is adopted at the national level. Better designs can be employed and conclusions then are likely to be less ambiguous, and political-institutional constraints are likely to be less severe. The intorduction of new programs can be staged so that earlier stages are a pilot for later ones. We stress formal tests of new program components and new variations here because such evaluations are not a matter of common practice. We will not learn how to bring about clear, detectable changes without more conscientious tests.

The second recommendation stems from our conclusion, based on this and other research, that better designs must be used if the Congress or the Department wants good estimates of the effects of programs on children. We do not advocate estimating those effects in all cases. The process is complicated under the best of conditions, despite cavalier announcements that the "program was successful because test scores went up" or that it was unsuccessful because they went down. We do advocate explicit authority in statutes for high quality designs, especially randomized experiments to facilitate their use. We believe explicit statutory provision is essential because such designs are the best in principle, and that should be recognized. The authorization should provide for review of the use of these designs.

Critique and Reanalysis of Evaluation Results

We recommend that in statutory requirements for evaluation of major programs, the Congress:

. also require independent, balanced, and competent critique of evaluation results that are material to policy decisions.

. require critique of samples of evaluations submitted by LEAs and SEAs in response to legal requirements

. require that statistical data produced by national evaluations be made available for reanalysis.

By critique here we do not mean adverse commentary. We mean reasoned judgments about whether conclusions drawn from the evaluation are sensible and can inform decisions. The main reason for the recommendation is that such criticism is not routine, but it is essential to enhance credibility of good evaluations, to properly identify poor evaluations as such, and to provide feedback to federal evaluation units, contractors, and grantees about the quality of their work. There is no formal system for competent critique of evaluation reports produced by LEAs and SEAs in response to law, yet many could benefit from criticism.

Use of Evaluation Results

We recommend that the Congress:

- direct staff of relevant committees, the Department, and the GAO to routinely outline which institutions can reasonably be expected to use results of each major evaluation and how such results might be used, during the design stage of every major program evaluation.
- specify exactly which evaluations have been used and why they were used, which have not been used and why they were not used, in authorizations and appropriations committee reports.
- require specific information about changes resulting from evaluation, whenever the law requires SEAs to describe uses of evaluation.
- explore the feasibility of direct competitive grants and contracts programs focused on improving the use of results at the LEA and SEA level.

The first recommendation's origins lie in the absence of any mechanism for planning use at the national level. Simply put, unless specific user groups are identified and some decision options laid out, evaluation results are less likely to be used. Indeed, if there is no clear way to link the evaluation with decisions or considerably better understanding, one can argue that the evaluation shouldn't be done at all. Specifying expectations will also help to make it easier to track utilization and that in turn will help to inform judgments about how evaluation resources could be better allocated. The recommendation to cite useful and useless evaluations in federal reports and to require SEAs and LEAs to record specific changes has the same objectives: understanding use better in the interest of better resource allocation. The suggestion to identify useless evaluation is not an invitation to criticize arbitrarily. We found that some LEAs and SEAs are capable and interested in inventing and testing better ways to use information. The suggestion to expand their opportunities for doing so is based on this.

Standards and Guidelines

Recently developed standards and guidelines for evaluation are not appropriate for incorporation into law. They are sufficiently well developed to recommend that the Congress:

- use such guidelines to understand what can reasonably be expected of evaluations.
- direct that agencies use them as a guide where appropriate to developing criteria for judging evaluation plans submitted by LEAs and SEAs.
- elicit assistance in the interpretation of guidelines from Congressional support agencies, such as GAO, that have been instrumental in their construction.

RECOMMENDATIONS TO THE DEPARTMENT OF EDUCATION

Authority for Technical Discussion

We recommend that the Department:

- authorize technical staff of evaluation units to initiate discussion of evaluation plans with pertinent Congressional staff, at their discretion, and refrain from directives which impede direct discussion.

The impetus for the recommendation is simple: Competent evaluators can expect to do a good job only when they have the opportunity to discuss Congress's information needs frequently. Restrictions on the evaluation unit's initiating discussion with Congressional staff of Committees that demand evaluation prevent the job from being done better. We recognize that some restrictions on bureaucratic lobbying for programs are warranted, and that some administrative rules are necessary to keep the process orderly. The lack of clear opportunity to figure out what Congress can use decreases the likelihood that evaluations will be timely, relevant, and credible, and the likelihood that the Congress will find the results useful. Relaxing restrictions will not of course guarantee usefulness.

Planning and Executing Evaluations

We recommend that the Department direct principal evaluation unit staff to meet regularly with relevant staff of committees to:

- negotiate agreement about when particular evaluations are warranted and the senses in which each evaluation required by law is possible.
- clarify Congressional information needs, quality of evidence required, and planning cycle for each major evaluation undertaken by the Department.
- identify specific audiences or groups for evaluation results.
- identify the changes in program or understanding which could occur on the basis of evaluation results.

The rationale for this recommendation is identical to the one offered for a similar recommendation made to Congress. Understanding Congressional information needs is not possible without some regular discussion between technical evaluation staff and Congressional staff. Scarcity of evaluation resources requires better planning and that planning cannot be informed without dialogue among relevant staff.

Tests of New Program Components, Program Variations, and New Programs

We recommend that the Department authorize explicitly the use of high quality evaluation designs, especially randomized experiments, in evaluating new program components, new program variations, and new programs, in all regulations which require estimating the effects of innovative changes.

The main justification is that high quality designs lead to far less debatable estimates of programs on children than low quality designs. They are more difficult to execute, and they are more feasible for pilot testing new programs, program variations, and program components, than for estimating the effects of ongoing programs. Explicit authorization would make the importance of good designs plain, and would provide more clear opportunity for competent SEAs and LEAs to exploit them.

Critique and Reanalysis of Evaluation Results

We recommend that the Department:

- provide for the independent, balanced, and competent critique of every major evaluation funded by the Department in procurement of evaluations and evaluation policy.
- incorporate into procurement procedures and policy the requirement that all statistical data produced in major program evaluations be documented and stored for reanalysis.
- create an administrative mechanism for deciding when simultaneous analysis by both the original evaluator and an independent analyst is desirable and feasible, and a mechanism for executing simultaneous independent analyses.

The rationale for this recommendation is identical to the one offered for a similar recommendation to Congress.

Access to and Specification of Reports

We recommend that the Department adopt a policy to:

- adhere to a clearance rule which makes evaluation reports available after a specified period of time.
- specify completely the evaluation documents referred to in the Department's Annual Evaluation Report, the Federal Register, and policy statements.

- include, in every major evaluation report, a list of core recipients of the report, or compiling publicly available lists of core recipients.

The recommendation stems partly from difficulties encountered in obtaining reports under review by the Executive Secretariate and other groups involved in the DHEW clearance process. We also found it difficult to identify reports precisely, when they were cited as evidence of the usefulness of evaluation in developing regulations or policy. The absence of a list of core recipients of reports makes it very difficult to identify potential user groups and to determine if reports were used. The consequence is that what is useless or useful is less verifiable.

The Use of Evaluation Results

We recommend that the Department direct evaluation unit staff or evaluation contractors to:

- provide oral reports regularly as well as written reports on results of major evaluations, and on the uses to which results can be put, to relevant Congressional staff and support agency staff and the program staff within the Department.

- create a system to periodically collect, synthesize, and report specific uses to which evaluations are put.

- improve the Annual Evaluation Report by citing instances of use more specifically.

- direct evaluation staff to meet regularly with Congressional staff to clarify information needs, feasibility of evaluation, audiences for results, and ways in which results can be used to modify programs.

The recommendations are based partly on the finding that use of evaluation results is not tracked conscientiously and the belief that it ought to be tracked to learn how to do evaluations better, and how to better allocate evaluation resources. The rationale for the last recommendation is identical to the one given earlier on planning and executing evaluations.

Implementation

We recommend that the Department:

- routinely require formal measurement of the degree to which program plans match actual operations.

- adjoin research on methods of measuring implementation to the introduction of new programs and program variations.

- create an inexpensive central information system on the time and resources required for full implementation of new programs.

The main reason for the first recommendation is simply that measurement of implmentation of innovations is infrequent. The reason for the second recommendation is that we know little about cheap effective methods of measurement in this arena. The third recommendation stems from the absence of any reasonable empirical guidelines on the time and resources necessary to implement innovative programs.

TABLE OF CONTENTS

1. Introduction..1- 1

    1.1 The Purpose of This Report..1- 1
    1.2 Project Staff: Qualifications and Independence......................1- 1
    1.3 The Research Startegy, Sources of Evidence, and the Orientation of the Project..........1- 2
    1.4 Organization of This Report...1- 4
    1.5 Preliminary Definitions...1- 4
    1.6 Agencies with Responsibility for Supporting and Conducting Evaluations and Developing Evaluation Methods............1- 5
    1.7 Elements of an Evaluation...1- 8

2. Why Are Evaluations Undertaken? To Whom Are Results Addressed?...........2- 1

    2.1 The Questions Addressed by Evaluation...............................2- 1
        . National Level: Illustrations.....................................2- 2
        . State Level: Illustrations..2- 3
        . Local Level: Illustrations..2- 4
        . Other Ways of Characterizing the Functions of Evaluation..........2- 7
        . Ambiguity in What is Meant by Evaluation..........................2- 8
    2.2 Audiences for Evaluation Results....................................2-10
        . National Level..2-10
        . The Local and State Interest in National Evaluations..............2-13
        . Local and State Audiences for Local Evaluations...................2-14
        . Consumer and Producer Education...................................2-16
    2.3 Legislation on Evaluation...2-17
        . Recency of Amendments and Enactments..............................2-17
        . Character of General Provisions...................................2-17
        . Specific Programs...2-18
        . Other Strategies of Evaluation....................................2-20
        . Types of Evaluation Activities Required by Education Statutes.....2-20
        . Types of Evaluation in Different Levels of Government.............2-23
    2.4 Functional Character of Evaluation: Statistical Description.........2-23
        . National Level..2-23
        . Special Studies...2-26
        . Local Level Evaluation..2-26
    2.5 The Decision to Evaluate..2-28
        . The Questions as Fundamental......................................2-28
        . Mechanisms for Deciding Whether to Evaluate.......................2-29
        . Evaluability Assessment...2-30
        . The Legislative Decision to Evaluate..............................2-31

3. How Are Evaluations Conducted?......................................3- 1

    3.1 Elements of an Evaluation...3 -1
        . Decision to Evaluate and the Questions to be Addressed............3- 1
        . Design of Evaluations...3- 2
        . Deciding Who Will Carry Out the Evaluation........................3- 2
        . Conduct of the Evaluation...3- 3

. Analysis, Reporting and Recommendations........................3- 3
. Evaluation of Evaluations......................................3- 4
. Dissemination and Use of Results...............................3- 4
. Obstacles..3- 4
. Evaluation Contexts: State Administered versus Direct Grants...3- 5
3.2 Procedures for Evaluation Specified by Law...................3- 5
3.3 Administrative Procedures: Federal Regulations...............3- 7
. Federal Reporting Requirements.................................3- 7
. Diversity in the Type of Evaluation Regulations................3-12
3.4 How Evaluations are Conducted at State and Local Levels......3-12
. State Level..3-12
. Types of State Education Agencies and Evaluation Practices.....3-13
. Illustrations of State Level Evaluations.......................3-14
. How Evaluations are Conducted at the Local Level...............3-21
. Existing Evidence on How Evaluations are Conducted Within LEAs.3-26
. Illustrations of Evaluation Practices at LEAs..................3-26
. Evaluation of Title I in Small and Large LEAs..................3-28
3.5 How Evaluations are Conducted at the Federal Level...........3-32
. Evaluation Within OED..3-33
. After the Decision to Evaluate: The Execution of an Evaluation.3-33
. Type of Questions Addressed and Characteristics of OED Studies.3-34
. Characteristics of OED Studies.................................3-35
. Evaluation in the Bureau of Education for the Handicapped......3-40

4. What are the Capabilities of Evaluators?......................4- 1

4.1 Misconceptions of Evaluators and Their Capabilities..........4- 1
4.2 Problems in Identifying Evaluators...........................4- 2
4.3 Specifying the Capabilities of Evaluators....................4- 3
4.4 The Organizational Factors Affecting Evaluation Activities and Capabilities....4- 5
. Federal Education Agencies.....................................4- 6
. State Education Agencies.......................................4- 7
. Local Education Agencies.......................................4-12
4.5 Capabilities for Evaluation..................................4-17
. Capabilities of Federal Education Agencies.....................4-17
. Capabilities of State Education Agencies.......................4-20
. Capabilities of Local Education Agencies.......................4-26
4.6 The Capabilities of Outside Contractors......................4-42
. The Capabilities of Outside Contractors Employed by Federal Agencies....4-42
. The Capabilities of Outside Contractors Employed by State and Local Education Agencies....4-44

5. How Well are Evaluations Carried Out?....................5- 1

5.1 Standards for Assaying Quality of Evaluation....................5- 1
. The Various Sources of Guidelines and Standards....................5- 1
. Differences and Similarities Among the Guidelines....................5- 2
. Commentary on and Experience with Guidelines....................5- 3
. Debate Over the Utility of Guidelines....................5- 3
5.2 Quality of Evaluations....................5- 3
. Previous Examinations of the Quality of Evaluations....................5- 5
. A Case Study of Three Contracted Evaluations....................5- 9
. Educational Standards for Success vs. Comparative Assessments....................5-11
. Impact of Regulations, Requirements and Legislative Mandates on Evaluation....................5-12
. The Evaluation Plan and the Proposal Review Process....................5-13
5.3 Critical Evaluation Issues....................5-16
. Program Implementation....................5-16
. Evaluation Design....................5-20
. Randomized Field Experiments....................5-22
. Quality of Analysis....................5-23
. Critique and Secondary Analysis....................5-24
. Storage of and Access to Information for Secondary Analysis....................5-24
5.4 Constraints on the Quality of Evaluation....................5-27
. Direct Contacts Between Congressional Staff and Agency Staff....................5-27
. Federal Clearance Requirements....................5-28
. Independence and Integrity....................5-29
. The Local Level....................5-31
. Constraints on Capabilities....................5-35

6. How are the Results of Evaluation Used?....................6- 1

6.1 Definition of use and Audiences for Evaluation Results....................6- 1
. Use of Information in Making Specific Decisions....................6- 1
. Use of Information to Enhance Understanding of Issues....................6- 2
. Use of Information to Persuade Others or to Confirm One's Beliefs....................6- 2
. Different Strokes for Different Folks....................6- 2
6.2 Sources of Data on the Uses of Evaluation Results....................6- 4
6.3 Use of Evaluation Results at the National Level....................6- 4
. The Annual Evaluation Report....................6- 4
. Congressional Staff Views....................6-10
. Congressional Use of Evaluations: Committee Reports....................6-11
6.4 State Uses of Evaluation....................6-19
. Federal Requirements That Use be Specified....................6-19
6.5 Local Uses of Evaluation Results....................6-21
. Site Visits....................6-21
. Use in Modifying Programs....................6-21
. Title I Programs and Test Scores....................6-24
. Routine Use in JDRP-NDN....................6-25
6.6 Understatement and Overstatement of Use....................6-27

6.7 Factors Influencing Use of Evaluations.............................6-30
6.8 Illustrative Case Studies.......................................6-34

7. Recommendations...7- 1

7.1 Recommendations to the Congress....................................7- 1
  . Planning and Executing Evaluations...............................7- 1
  . Statutory Provisions for Evaluation.............................7- 1
  . Evaluation Capabilities...7- 2
  . Use of and Authority for Better Evaluation Designs...............7- 2
  . Critique and Reanalysis of Evaluation Results...................7- 3
  . Use of Evaluation Results.......................................7- 4
  . Standards and Guidelines..7- 4
7.2 Recommendations to the Department..................................7- 5
  . Authority for Technical Discussion...............................7- 5
  . Planning and Executing Evaluations...............................7- 5
  . The Use of Evaluation Results...................................7- 6
  . Implementation..7- 6
  . Tests of New Program Components, Program Variations, and New Programs.....7- 7
  . Critique and Secondary Analysis.................................7- 7
  . Access to and Specification of Reports...........................7- 7
7.3 Rationale..7- 9
  . Planning and Executing Evaluations...............................7- 9
  . Statutory Provisions for Evaluation.............................7-11
  . Authority for Using Better Evaluation Designs....................7-12
  . Independent Critique and Secondary Analysis......................7-13
  . Access to Reports...7-16
  . Tracking the Use of Evaluations.................................7-17
  . Standards and Guidelines..7-20
  . Estimating the Effects of Programs...............................7-22
  . Facilitating Integrity..7-24
  . Evaluation Capabilities...7-27

8. References

Appendices

CHAPTER 1. INTRODUCTION

Robert F. Boruch and David S. Cordray

Mandates come in as many varieties as
finches, warblers, sparrows and...they
are generally divided into the Greater
Mandates and the Lesser Mandates.

Eugene J. McCarthy and James J. Kilpatrick
A Political Bestiary

1. INTRODUCTION

1.1 THE PURPOSE OF THIS REPORT

This report concerns evaluation of federally supported educational programs at the national, state, and local levels. It was undertaken in response to Section 1526 of the Education Amendments of 1978 (Public Law 95-561), which requires that the Commissioner of Education conduct a comprehensive study of evaluation practices and procedures. The questions covered here are those implied by the law and the conference reports preceding it, and those enumerated in the Work Statement for this project:

- Why and how are evaluations carried out?
- What are the capabilities of those who carry out evaluations?
- How are the results of evaluation used?
- What recommentations can be made to improve procedure or practice?

We discussed the questions with Congressional staff and federal agency personnel to clarify them. The more detailed questions are elaborated in the body of this report.

The Project on Evaluation of Evaluations is prospective in its orientation, designed to provide evidence and argument bearing on these questions and to provide recommendations which will help to ameliorate the problems we have identified. Pertinent excerpts from the law, conference report, and Work Statement for the Project are given in Appendix 1 to this report.

1.2 PROJECT STAFF: QUALIFICATIONS AND INDEPENDENCE

Northwestern University was given responsibility for conducting this study. Staff members for the Project were drawn primarily from the University, and their efforts were supplemented by consultants from universities, education agencies at the local and state levels, and private institutions. The staff members were selected so as to include individuals with expertise in evaluation methods and policy, education, law, management, and psychology. A description of staff members, consultants and their roles, is given in the Appendix of this report.

Neither of the principal investigators in this study, Boruch and Cordray, bid on competitive evaluation contracts awarded by the U. S. Office of Education, the National Institute of Education, or any other agency. This appraisal then is administratively and fiscally independent of school districts, states, the federal government, private vendors, and other regular contractors. We are not independent of federal agencies in the sense of having received grants from the National Science Foundation and NIE for research, and having participated in advisory boards to other agencies with and without payment. To assure that other independent appraisal of evidence in this report is possible, we furnish references to published and unpublished work. Unpublished material which does not abridge privacy of individuals is stored at Northwestern University and will be made available to other analysts so long as resources permit.

1.3 THE RESEARCH STRATEGY, SOURCES OF EVIDENCE, AND THE ORIENTATION OF THE PROJECT

To answer the questions posed earlier, the Project staff relied on two broad sources of information: contemporary research and development by other researchers, and direct investigations by the Project staff. Round-table discussions were organized to enlarge on information obtained from these sources. In considering contemporary research we reviewed over 400 recent published and unpublished reports on functions, conduct, and staff of evaluations at local, state, and federal levels, and pertinent testimony to Congress. Raw data were acquired and reanalyzed when this was warranted and time was sufficient to assay the data's quality. This review of others' work was supplemented by interviews with the individuals responsible for the production of some of the reports, such as government staff and contractors. The results of the review are employed throughout the report. A consolidated review of the work is presented in the Appendix and our review of evaluation policy has been published in Review of Research in Education.

The second major resource was information collected by the Project staff directly from individuals on site, in telephone interviews, and through round-table discussions. The purposes of this exercise were to assure understanding of earlier work and to fill in gaps in what was known about evaluations when the Project began. Such interviews were essential in corroborating evidence about the use of evaluations, covered in Chapter 6, for example. At the federal level, the interviews focused on staff of the Office of Evaluation and Dissemination at the U.S. Office of Education. Interviews were also conducted with members of the National Institute of Education, pertinent Congressional staff, the General Accounting Office, the Congressional Budget Office and the Congressional Research Service. The main criteria for selecting respondents in each of these cases were their knowledgeability about educational evaluation.

At the state level, site visits were made to six state education agencies: Minnesota, Michigan, California, Texas, New Jersey and Massachusetts. They were selected randomly from a stratified list of sites. Florida was chosen purposively for a site visit on account of Florida's use of particular evaluation approaches in Title I supported compensatory education programs. Columbus, Ohio, served as a pilot test site. Approximately 50 telephone interviews were undertaken in the remaining states to supplement site visits with statistical characterization.

At the local level, twelve school districts were selected randomly for site visits and intensive case study from a stratified list. Three districts that declined were replaced by three selected randomly from a matching list. The resulting sample of cases includes:

Jersey City, New Jersey
San Diego, California
Amarillo, Texas
Gaston County, North Carolina
Broward County, Florida
St. Louis, Missouri
Springfield, Massachusetts
Lansing, Michigan
Colorado Springs, Colorado
Kanawha County, West Virginia
St. Paul, Minnesota
Jefferson Parish, Louisiana

In addition, Providence, Rhode Island, Marion County and Osceola County, Florida, were visited to assay their use of more sophisticated approaches to evaluating Title I programs. The individuals to whom we spoke in each visit included administrators, program staff and teachers, school board members, and evaluators. The case studies were supplemented by over 200 telephone interviews with individuals having evaluation responsibility at the local level. The sample was selected randomly from a stratified list of LEAs. The design is described in Appendix 3 of this report.

Private and public organizations serve as contractors for federally supported evaluations and for some field research on evaluation. Individuals who cooperated in supplying information, corroborative evidence, or reactions to the questions we were asked to address were staff members of: Huron Institute, Abt Associates, Bay Area Research, NTS Research Corporation, Bureau for Social Science Research, Systems Development Corporation, Westat, Rand Corporation, Educational Testing Service, state departments of education, local research and evaluation units, and others mentioned in the text.

Round-table discussions were organized to consider particular topics and participants were invited on the basis of their expertise. The topics included "School Boards and Evaluation," "Evaluator Capabilities," "Utilization," and "Vocational Education." Participants included representatives from school districts such as Evanston and Chicago (Illinois), Mesa (Arizona), Austin (Texas), states such as Minnesota and California, research organizations mentioned earlier, and independent university researchers. Round-table participants were treated as consultants.

Our time schedule did not permit presenting more than fragments of our report in professional forums. We have been able to capitalize, however, on brief presentations made at the annual meeting of two national professional organizations, at meetings of the National Academy of Sciences Committee on Program Evaluation, and at professional meetings on evaluation organized by the Department of Justice and the National Institute of Education.

This Project formally began September 29, 1979, and submission of a final report was scheduled for June 30, 1980. Because of this tight schedule, we focused our attention primarily on evaluations in four program areas: compensatory education supported under Title I, education for the handicapped, vocational education, and bilingual education. These served only as general targets and we capitalized on some work in other areas, notably career education, day care education, innovative projects, and higher education. Our early discussions with federal agency staff and Congressional staff led to an agreement that our findings and recommendations should, however, be directed to evaluation generally rather than evaluation in specific program areas. We have abided by that agreement generally in this report.

The general orientation of the Project is prospective as agreed during meetings with Representative Holtzman's staff. This study has not been dedicated to identifying evaluations or evaluation staff that are particularly inept. Rather, the aim has been to develop useful general statements and to make recommendations based on available evidence, rather than to publicly criticize particular individuals. Some recommendations are, however, dedicated to reducing the likelihood of incompetence and to facilitating integrity. The information we have collected on individuals is understood to be confidential and for research purposes alone, helpful in making decisions about quality and use of evaluation, and not for making decisions about specific individuals.

A detailed description of surveys, case study selection, methods roundtable participants, and other technical details on information sources is given in Appendix 3.

1.4 ORGANIZATION OF THIS REPORT

The report is organized around the major questions we were asked to address. Chapter 2 considers the rationale, evidence, and opinion bearing on why evaluations are done, the confusion and argument engendered by general demands for evaluation, and the audiences to whom evaluations are addressed. Chapter 3 addresses the question of how evaluations are executed. Chapter 4 covers the organization of evaluations and the capabilities of evaluators, and Chapter 5 considers quality of evaluations. The way evaluation results are used is considered in Chapter 6. The chapter includes case studies on the use of evaluative information. Chapter 7 covers recommendations.

1.5 PRELIMINARY DEFINITIONS

Section 1526 of Public Law 95-561 refers to "evaluations at federal, state, and local levels." To avoid confusion, we exploit existing federal agency guidelines and recognize that there is no universally accepted definition of evaluation. Consequently, we adopt a working definition and catalog a set of questions often addressed in evaluation. Evaluation is defined here tentatively as a study designed to answer questions about what a program does in the interest of making judgments about the program. The questions often addressed include: Who is served? What services are delivered? At what cost? With what effect? A more elaborate description is given in Chapter 2, along with a discussion of the diverse meanings attached to the word by academicians, legislators, program managers, and others.

Local education agency (LEA) is defined as a school district or organizational unit within the school district. State education agency (SEA) is defined as a state organization responsible for administering any educational program at the primary or secondary school level. Local and state level evaluation refer to activities at each level. References to federal agency will normally specify which agency is pertinent. Contractor here refers to any public or private organization funded to design or conduct an evaluation, or to provide technical assistance through training, analysis, and so on.

The Work Statement for this Project refers to proposals and recommendations for revision or modification of evaluation practices and procedures. We define this further to mean suggestions of the following kinds:

- clarification of language on practices and procedures,
- enumeration of options on practices and procedures,
- specification of principles which should underlie practice
- specification of direct implications of contemporary evaluations for practice and procedures,
- establishing priorities,
- specification of action or decision.

1.6 AGENCIES WITH RESPONSIBILITY FOR SUPPORTING AND CONDUCTING EVALUATIONS AND DEVELOPING EVALUATION METHODS

The U.S. Office of Education's Office of Evaluation and Dissemination (OED) has had primary responsibility for evaluation of OE's education programs. The main focus of this Project has been on OED activity. Other agencies, such as the National Institute of Education, do undertake evaluative work at times. Oversight agencies, such as the General Accounting Office have had responsibility for activities which can be properly labelled as evaluation as well as for overseeing evaluations undertaken by other federal agencies. The following remarks briefly describe the pertinent agencies and the scope of their work. The recent changes engendered by creation of the new Department of Education are noted and an organization chart for the Department is attached.

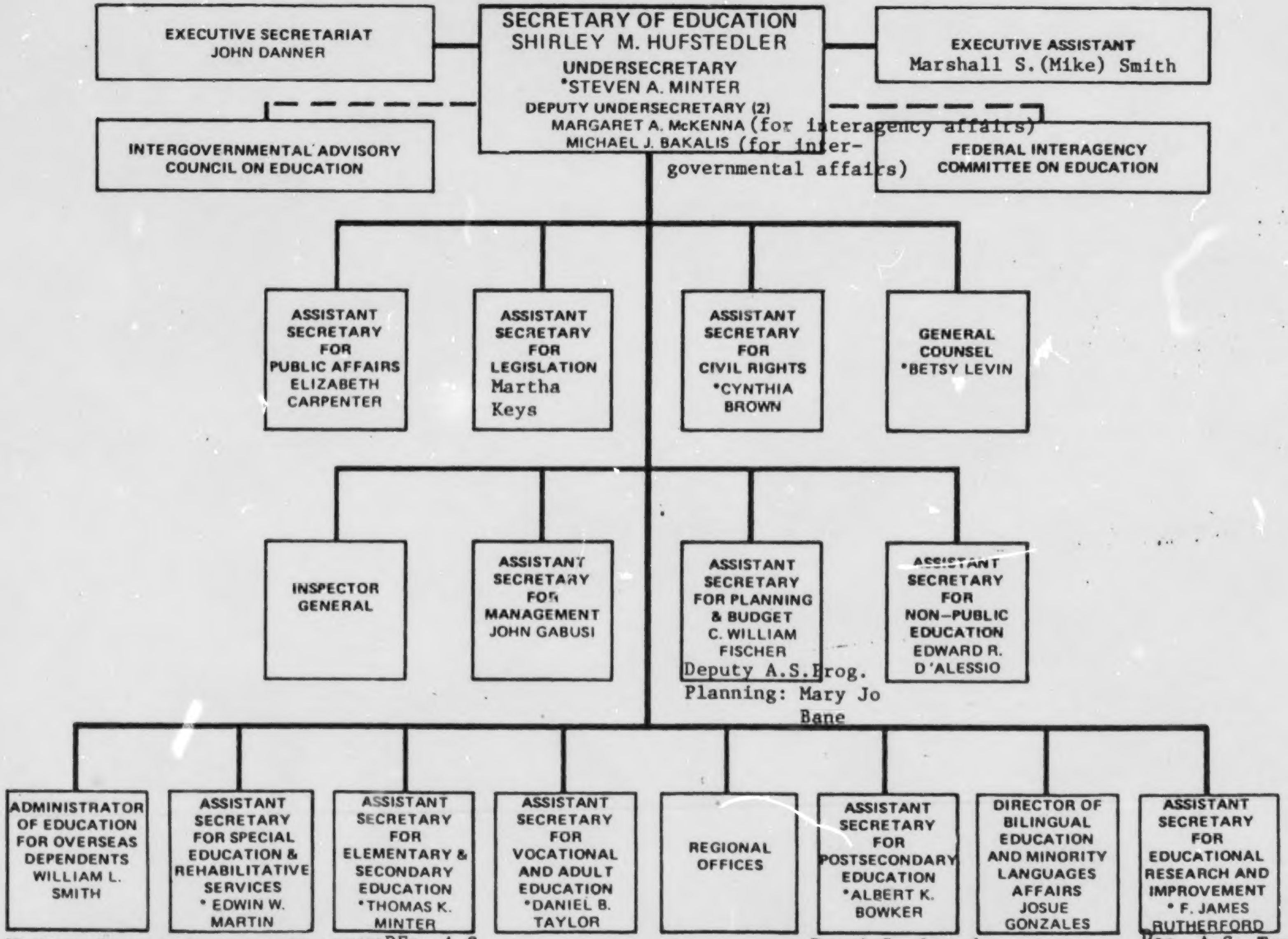
Over the past five years, The Office of Evaluation and Dissemination has been responsible for evaluating programs administered by the U. S. Office of Education and for coordinating dissemination of exemplary materials to state and local education agencies. The major routine exceptions to the evaluation mission have been the Bureau of Education for the Handicapped, the Bureau of Student Financial Aid, and the Follow Through Program, each of which administers its own evaluation funds. The Office's predecessor, created in 1970, has undergone changes in title, such as the Office of Program Budgeting and Evaluation, and in responsibilities. As of March 1980, the Office had 40 professional staff members. OED has been responsible for administration of an average of $21 million annually in evaluation funds since 1975. Its evaluation responsibilities have included deciding whether to evaluate, designing evaluations, issuing evaluation contracts, monitoring, reviewing, and criticizing evaluations which it supports. OED routinely prepares summaries of evaluation for dissemination and has routinely distributed both summaries and reports. The evaluation unit of OED has been divided into three units, each being responsible for a specific education sector: Elementary and Secondary Education; Occupational, Handicapped, and Development Programs; Post-secondary Education. A fourth Division of Educational Replication was responsible for dissemination policy and disseminates educational products approved by the Joint Dissemination Review Panel (JDRP). JDRP has included staff members of the Office of Education and the National Institute of Education. OED released its ninth Annual Evaluation Report covering fiscal year 1979.

Under the new Department of Education, the evaluation staff of OED were transferred to a new unit within the Department - the Office of the Deputy Assistant Secretary for Evaluation and Program Management, headed by John Seal. Program Evaluation is one of three divisions within the unit. The remaining two are the Division of Management Evaluation and the Division of Management Planning and Assistance. The Evaluation Division, according to internal policy statements, has the responsibility to conduct impact evaluations and formative or process evaluations, and assess alternative program strategies and structures. The Division will continue to prepare requests for Proposals and review and approve evaluation contract proposals. The Dissemination staff of OED have been transferred to the Office of Dissemination and Professional Improvement in the new Office of the Assistant Secretary for Educational Research and Improvement.

The Education Amendments of 1972 (Public Law 92-318) established the National Institute of Education as part of the Education Division of the Department of Health, Education, and Welfare. NIE is charged under its enabling statute with building an effective research and development system in the interest of improving American education. This includes administrative and fiscal responsibility for educational laboratories and centers, such as UCLA's Center for the Study of Evaluation, and for the National Assessment of Education Progress run by the Education Commission of the States. The NIE acts as a foundation in making grants to independent researchers in universities, colleges, educational organizations, state and local educational research units and awards contracts for larger scale, applied research on special research topics.

Evaluation-related work includes development and testing of new methods of evaluation, new solutions to problems of data access and dissemination, new methods of testing and observation, and development of guidelines on evaluation for use by local and state education agencies. The methodological work is administered by NIE's Testing, Assessment, and Evaluation Division and is carried out by the Northwest Regional Education Laboratory, for instance, as well as by independent researchers. NIE has no routine responsibility for evaluation of programs administered by the U. S. Office of Education. But evaluation of smaller new pilot programs whose development is supported by NIE falls within its research mission. NIE has at times

U.S. DEPARTMENT OF EDUCATION


EXECUTIVE SECRETARIAT JOHN DANNER
SECRETARY OF EDUCATION SHIRLEY M. HUFSTEDLER
UNDERSECRETARY *STEVEN A. MINTER
DEPUTY UNDERSECRETARY (2)
MARGARET A. McKENNA (for interagency affairs)
MICHAEL J. BAKALIS (for inter-governmental affairs)
EXECUTIVE ASSISTANT Marshall S.(Mike) Smith
INTERGOVERNMENTAL ADVISORY COUNCIL ON EDUCATION
FEDERAL INTERAGENCY COMMITTEE ON EDUCATION
ASSISTANT SECRETARY FOR PUBLIC AFFAIRS ELIZABETH CARPENTER
ASSISTANT SECRETARY FOR LEGISLATION Martha Keys
ASSISTANT SECRETARY FOR CIVIL RIGHTS *CYNTHIA BROWN
GENERAL COUNSEL *BETSY LEVIN
INSPECTOR GENERAL
ASSISTANT SECRETARY FOR MANAGEMENT JOHN GABUSI
ASSISTANT SECRETARY FOR PLANNING & BUDGET C. WILLIAM FISCHER
Deputy A.S.Prog. Planning: Mary Jo Bane
ASSISTANT SECRETARY FOR NON-PUBLIC EDUCATION EDWARD R. D'ALESSIO
ADMINISTRATOR OF EDUCATION FOR OVERSEAS DEPENDENTS WILLIAM L. SMITH
ASSISTANT SECRETARY FOR SPECIAL EDUCATION & REHABILITATIVE SERVICES * EDWIN W. MARTIN
ASSISTANT SECRETARY FOR ELEMENTARY & SECONDARY EDUCATION *THOMAS K. MINTER
DEp. A.S. Cora Beebe
ASSISTANT SECRETARY FOR VOCATIONAL AND ADULT EDUCATION *DANIEL B. TAYLOR
REGIONAL OFFICES
ASSISTANT SECRETARY FOR POSTSECONDARY EDUCATION *ALBERT K. BOWKER
Dep.A.S. Joseph Shenker
DIRECTOR OF BILINGUAL EDUCATION AND MINORITY LANGUAGES AFFAIRS JOSUE GONZALES
ASSISTANT SECRETARY FOR EDUCATIONAL RESEARCH AND IMPROVEMENT * F. JAMES RUTHERFORD
Dep. A.S. Terry Saario

**Pending confirmation by the Senate as of May 1, 1980*

GPO 887-878

been given a special Congressional mandate to evaluate, e.g., the Compensatory Education Study of Title I supported programs and the current Study of Vocational Education. And it has been directed occasionally by the Secretary of DHEW to undertake special evaluations. Current efforts include assessments of the Reverend Jesse Jackson's Push for Excellence Program and the Cities in School program. In the new Department of Education, NIE is a part of the Office of the Assistant Secretary for Educational Research and Improvement.

The Bureau of Education for the Handicapped has had a routine responsibility for small scale evaluations of projects for the handicapped as part of its development mission. The vehicle for this activity is BEH's Division of Innovation and Development. A unique major evaluation activity during 1977-78 has been developing evaluation plans for Public Law 94-142, a statute mandating free and appropriate public education for all handicapped children. Primary responsibility for the plan and its execution has been lodged with the State Program Implementation Studies Branch. The evaluation focused on questions similar to those typically addressed in studies run by other agencies: Who is served? What are the services? To what extent is the intent of the Act met? And so on.

The Office Assistant Secretary for Planning and Evaluation in DHEW's Education Division has had responsibility for review and synthesis of evaluations executed by other agencies, and using of evaluations in planning. The number of actual evaluations carried out since 1978 has been small; most activities focus on planning. The Office of the Assistant Secretary in DHEW's Evaluation and Technical Analysis Division has responsibility for small evaluability studies - examining the extent to which program objectives are measurable.

The National Center for Educational Statistics is mandated under the law to "collect and disseminate statistics and other data related to education in the United States and in other nations." NCES's activity is confined to descriptive surveys rather than to evaluations, but the information generated is material to design and execution of evaluations. That information includes, for instance, listings of school districts and other educational institutions and their characteristics, and other information which is helpful in designing special purpose evaluations. More generally, the descriptive statistics generated in surveys, on expenditures, pupil characteristics, enrollments, and the like normally serve as backdrop for educational research and evaluations. Some data collection efforts, notably the National Longitudinal Study of the High School Class of 1972 and of class of 1980 serve as the empirical basis for theoretical analyses of the process and outcomes of education. This includes efforts by academic researchers to estimate the effects of federal programs based on the survey data rather than controlled field evaluations.

Under Title VII of the Congressional Budget and Impoundment Control Act of 1974 (Public Law 93-344), the Comptroller General is required to "review and evaluate the results of Government programs and activities. . .

when ordered by either House of Congress, or upon his own initiative, or when requested by any committee. . .or joint committee. . .having jurisdiction." In implementing the law, the General Accounting Office created a new Program Analysis Division to develop perspective, policy, and guidelines on evaluation and other approaches to understanding performance of federal programs. The Human Resources Division has had major though not exclusive responsibility for assessments of educational programs. In 1976, for instance, it reported on problems and needed improvements in evaluating education programs. Current efforts include developing a report on the quality of evaluations. The evaluation-related activities of PAD cover a wide range including development of status and issues papers on the topic, development of guidelines for assessing quality of evaluation reports, and guidelines and policy for reconciling privacy problems engendered by federally supported social research in general. In 1980, a new Institute for Program Evaluation was set up within GAO partly to consolidate initiatives which cut across divisions such as PAD and Finance and Management Sciences. The 1974 mandate and subsequent activity reflect a notable departure from earlier roles of the GAO, going well beyond the management and accounting emphasis of the 1950's and bringing a wider variety of skills and interests into the organization.

The National Science Foundation's primary mission is support of scientific research. NSF has been involved in the development of science education programs and to a limited extent in their evaluation. The agency has also supported basic and applied research on methods which ultimately find their way into evaluations in education, health, economics, criminal justice and law enforcement and other areas. This includes, for instance, support of the development of a state of the art work on the use of formal field experiments to plan and evaluate social programs, and on solutions to managerial, scientific, and political institutional problems engendered by such field tests.

State and Local Education Agencies with Responsibility for Supporting or Conducting Evaluations

At the state level, organizations responsible for evaluation differ from one education agency to the next, and from state to state. No single organizational entity is responsible for evaluation. Within a school district, responsibility for evaluation may be organized along program lines, a Title I program manager being responsible for evaluation, for example. Or, the responsibility may be vested in a research and evaluation unit. The responsibilities imposed by federal law are discussed generally in Chapter 2. Organizational arrangements are considered in the chapter on manpower capabilities.

1.7 ELEMENTS OF AN EVALUATION

The word evaluation implies different things to different people. To avoid some chronic misunderstandings here, we enumerate the elements here briefly. The elements are, in principle, desirable judging from guidelines issued by professional organizations, by Congressional support agencies such as the General Accounting Office, and by federal evaluation agencies. They are not always a matter of practice:

- Deciding to evaluate and choosing the questions to be addressed by evaluation
- Designing the evaluation, including sample design
- Contracting for the evaluation
- Conducting the evaluation and pertinent side studies
- Analyzing results, making recommendations, and reporting
- Using the results

The elements are discussed in Chapter 3, on how evaluations are conducted.

CHAPTER 2. WHY ARE EVALUATIONS UNDERTAKEN?

Robert F. Boruch, David S. Cordray, Joe S. Cecil, and Laura Leviton

Program Evaluator: Democritus said that he would rather discover a single causal connection than sit on the throne of Persia.

Program Manager: Some of us would rather sit on the throne of Persia.

2. WHY ARE EVALUATIONS UNDERTAKEN? TO WHOM ARE RESULTS ADDRESSED?

This chapter discusses why evaluations are carried out and identifies some of the audiences for evaluation results. The most direct reasons for evaluation are considered in the next section with short illustrations from local, state, and federal experience. The justification for evaluating is complicated by ambiguity in the use of the word evaluation and we also consider this topic. The audiences for evaluation, the groups for whom evaluations are produced and those exhibiting interest in evaluation results are discussed in 2.2. We present information on statutory mandates for evaluation in Section 2.3 because the laws are an immediate justification for evaluation, but certainly not the only one, at all levels of government. Section 2.4 presents statistical information about the questions addressed by local, state, and federal level evaluations. Section 2.5 reviews issues in the decision to evaluate.

2.1 THE QUESTIONS ADDRESSED BY EVALUATION

In principle, most evaluations are carried out to answer one or more of the following questions:

Who is served by the program? Who needs services?
What are the services, how well are they delivered, and what do they cost?
What are the effects of services on recipients?
What are costs and benefits of alternatives?

Moreover, the information is obtained to facilitate making judgments or decisions about some aspect of the program. The audiences for the information depend partly on which questions are answered and may include policy makers, managers, and oversight groups.

This description is, of course, deceptively simple. Matters become complicated quickly once the decision to evaluate is made. So, for instance, the Congressionally mandated Compensatory Education Study of Title I programs focused on only one audience, Congress, deciding that earlier efforts to accommodate all possible audiences, such as interest groups and federal program managers, were inappropriate. That decision engendered problems in dealing with audiences whose interests were not addressed directly. The Study was remarkable in being asked to address questions about fundamental purposes of the program, along with more general questions of the sort enumerated earlier. New questions were adjoined to the ones specified originally by Congress so as to satisfy special concerns. The questions themselves had to be translated into rather more specific form to be useful. The translation resulted in the Study's focusing on four broad topics--allocation and distribution of funds, relations among government agencies in regulating and managing program activities, delivery of services, and changes in abilities of students participating in the program. The equal priority attached to each of these themes ran counter to then prevailing opinions which then attached highest priority to estimating effects on children. The latter was found not consistent with the Congress's view that other aspects of the program were equally important. The process of negotiation and approval on questions to be addressed, on the strategy to be used in answering them, in anticipating how the information would be used took over six months.

There are, of course, a variety of reasons why evaluative questions are asked in the first instance. We cannot deal with all of them here, but three of the most salient are worth mentioning. The fundamental one concerns children. Programs that are evaluated are dedicated to aspects of children's well-being which can be influenced by schools. To the extent that evaluation helps one understand whether a child is served, how well he or she is served, and whether beneficial effects are detectable, then evaluation is part of the program and seeks the same ends. The second reason is simple accountability. To the extent that programs are directed toward special needs and they are expensive in the short run, it makes sense to establish that they are not diverted in unproductive ways, to determine if they warrant improvement. The third reason bears on compromise. This includes, for example, suggesting that pilot tests of a program, an evaluation of a special kind, be undertaken before a massive new program is mounted, in the face of vigorous but insufficiently informed enthusiasm. The issue of the new program then is shelved until more information is obtained.

National Level: Illustrations

The evaluation strategy of the U.S. Office of Education has been pertinent to the questions enumerated above. Elaborated in USOE's Annual Evaluation Report, the strategy includes clarifying feasible goals for programs, identifying modifications in program content or administration to improve programs, and identifying especially effective projects. These core questions are also explicit in the GAO's guidelines on impact evaluation. Judging from interviews with CBO staff, the questions are typically those addressed by staff members in education as well.

Answers to questions about who received compensatory education services under Title I, for example, have been provided most recently by the Sustaining Effects Study and by the National Institute of Education Compensatory Education Study.[1] That head counts such as these are deceptively simple is apparent from, for example, the discrepancy between estimates yielded by state reports of students served (5 million) and estimates yielded by the NIE Compensatory Education Survey (6 million). Determining who receives and who might receive services under Public Law 94-142 for the handicapped was undertaken by the Bureau of Education for the Handicapped. This effort was remarkable in finding a major discrepancy between estimates of the number of handicapped children provided in the law, eight million, according to Section 601 of U.S.C. 1401, and estimates based on field research--4 million in 1977-78. This, of course, has major implications for appropriations.

Questions about who receives Title I federal monies for services, at what level, and of what kind have been addressed in the early reports of the Sustaining Effects Study supported by OED. The work focused on alternate ways of defining children's eligibility and institutional eligibility, and the probable effect of changes in rules. The actual operation of certain programs was the focus of Rand's study of programs supported under Title IV of the Civil Rights Act and dedicated partly to resolving education problems engendered by school desegregation.

Efforts to estimate the effect of programs on children are not undertaken often, partly on account of the resources required to mount high quality tests. Among early efforts, we include controlled field experiments such as the one conducted by the National Opinion Research Center and dedicated to estimating the relative effects of Emergency School Assistance Act programs in facilitating education in schools undergoing desegregation. The more recent outcome evaluations include efforts to estimate relative effects of compensatory education programs such as Follow Through, of bilingual education, and of Title I. The analysis and the ensuing debate over interpretation over what one could infer from the Follow Through evaluation illustrate the difficulty of such outcome evaluations. The matter is considered later in this report.

State Level: Illustrations

At the state level, answers to the question "Why evaluate?" are limited by the fact that many states take as their responsibility technical assistance and coordination rather than actual evaluation. The immediate reason for any involvement is federal law. But some states have taken a strong initiative to develop sophisticated approaches that are consistent with federal demands and exceed them. One of the typical incentives for this is that some states have their own programs running in parallel to federally supported programs. State education agencies are normally responsible under the law for disbursement of funds provided under Title I and for reporting and periodic monitoring of local programs supported by Title I. The reporting on "outcomes" often consists of consolidation of achievement test results supplied by local education agencies. Where the states have a reporting role in compensatory education, in vocational education, and others, reporting focuses on consolidating information about who is served, the nature of services, and efficiency in delivery.

Some states exceed federal requirements in having developed remarkable evaluation reporting systems. Minnesota, for instance, requires specification of objectives and evaluation plans for education programs, and Minnesota law on educational planning and evaluation is distinctive. Massachusetts law requires that local education agencies specify clearly how Title I programs are modified on the basis of each evaluation. California incorporates requirements for evaluation into its Comprehensive

School Improvement Program (Assembly Bill 65, Chapter 894, 1977), and the state's Master Plan for Special Education (Assembly Bill, 1250, Chapter 1247, 1977), and others.

Local Level: Illustrations

From our site visits to local education agencies, we judge that government requirements are an immediate reason for evaluation of all programs receiving federal support. But this does not mean that all LEAs merely comply with state or federal requirements. There is also considerable variation across LEAs and across programs within the LEA.

Variation across Programs. To be specific, we asked our respondents directly why evaluations were undertaken and the priority they attached to federal and local demands. For Title I programs, about half the sites we visited put local interests in evaluation far above government requirements. The remaining half do evaluations primarily because they are required. There may be argument about what it means to "find out how well the program is working..." in the more active sites, in Title I and other programs. But that debate does help to illuminate the evaluation and provoke interest. For bilingual education, interest in using evaluations to modify programs takes a slightly higher priority than federal requirements in our site visits. In vocational education, the majority of sites "evaluate" mainly to meet federal requirements. The importance of those requirements is overshadowed by local interest in evaluation in a minority of cases. Innovative projects supported under Title IV-C include vocational education and the interest in evaluation apart from meeting requirements is clear. In programmatic education for the handicapped, the stress is on meeting government requirements. It is a relatively new program, the idea of evaluation apart from compliance is not well developed, and sites engaged in little or no systematic evaluation beyond this.

Variation across Sites. There is considerable legitimate interest in finding out whether a program is working in the active sites and programs. To be sure, there may be argument about what "working" means. Some regard it as an easy question and simply obtain counts of those who are served. Others try to estimate unambiguously the effect of services. But the interest is explicit at local and state levels which go beyond government requirements.

To illustrate, Site A's reports on evaluation of an Emergency School Assistance Act program went considerably beyond requirements in attempting to estimate the effect of the program on children, teachers, and parents. Site D regularly augments the data collected to meet Title I requirements and uses the augmented information to assure that program objectives are appropriate rather than gratuitous and to set annual student achievement goals. Questions bearing on the effect of programs are not accorded high priority because staff believe it's not possible to estimate the effect independent of other services to children, in the case of Title I, or because the program has already "demonstrated its effectiveness." Site E's commitment to an office of research and evaluation stemmed from early federal mandates to evaluate. The district has increased the office's evaluation budget regularly, and this appears to be a result of energetic production of balanced reports and an interested, responsive school board and administration.

Not part of the site visit sample, but no less interesting, are LEAs whose staff members have published accounts of their activity. The Detroit Public School District, for instance, has cooperated with Wayne State University staff to mount well designed randomized field experiments to determine the effects of a preventive mental health program for children at risk and enrolled in Title I programs. McDuffie County, Georgia's LEA has had sufficient interest and resources to develop a Title I program and an evaluation which are sufficiently sturdy to pass muster with the Joint Dissemination Review Panel, and the program is now being made available through the National Diffusion Network. About 45 Title I programs have been recognized by JDRP. Providence, R.I., is remarkable for attending to a wide range of problems encountered by Title I programs and being able to document evaluations of tests, program goals, program components, staff concerns, parental involvement and other matters.

On the other hand, Site K generally does "evaluation" only because it's required, for example, and this amounts to no more than simple reporting of test scores. In Site J, the evaluation of the special education program for the handicapped amounts to no more than setting an objective, such as "getting a typewriter" and achieving the objective "got the typewriter." Neither the school board nor administration appear to be much interested in evaluation in Sites A, C, H, J, and K and that disinterest is reflected at the program level. Those sites which stress meeting federal requirements are not all uninterested in evaluation. Some respondents told us they would like to do more, but local indifference or lack of time and money prevent doing more. Some respondents clearly saw no point to doing anything more than the government required.

Reasons for the Questions. The immediate reasons for addressing the questions at the local level is, as we've said, federal requirements. But other reasons are as numerous as those at any other level of government. Where school boards are active and interested in evaluation, evaluators are accountable to their members. Where superintendents are vigorous in supporting an evaluation unit, their interest stems partly from routine management information which is essential in operations and some of the information needed to modify operations. The benefits of evaluation to children emerged indirectly in conversations partly because there is some reluctance among evaluators to announce that ultimately evaluation is "for the children." The phrase is used hypocritically as well as honestly in the field, and the chronic hypocritical uses make honest uses something of an embarrassment. To illustrate the benefits, consider Site E's recent tests, designed to understand what strategies facilitate children's reading. An expensive, commercially advertised regimen was compared against an in-school remedial teaching program and against one other approach to assay how children performed. The results demonstrated that children learned more under the regimen only when it was exceedingly well implemented. Since high implementation was difficult, and since the alternatives did not fare well relative to the new regimen, the results were used to argue that districts should be allowed to choose between the remaining ones.

Apart from interest in answering questions, a common reason for evaluation is verifying to others the worth of a program. The Title I director in one site, for instance, stressed the fact that he wanted to assure that if he did a good job, there could be an independent appraisal of the effort. His interest lies in building a small pocket of credibility in a city which has a fairly strong tradition of corruption. We encountered sufficient recognition of the benefit of independent verification at other sites to believe that this reason is common and neither demeaning nor dishonest.

We found no uniform attention across programs in estimating the effects of programs on children. Indeed, there was informed skepticism in Sites A, D, J, and others that estimates of the effects of Title I services could be made sensibly, even with federal "models" of how to do it, because other services were provided with Title I services simultaneously. In other cases, the task of estimation is easier simply because there are no other special services. The skepticism is reflected at the federal level as well judging from professional papers published by staff members of ASPE, academic critics, and others. We have not seen the same skepticism registered by LEA or SEA officials in testimony to the Congress. The difficulty of estimating effects generally has not prevented small scale tests from being mounted where the opportunity arises. In Site J, controlled tests of a vocational education program have been nicely designed and executed using Title IV-C funds for innovative projects. Recall the Detroit and McDuffie County illustrations given earlier. The remarkable efforts are in a minority to be sure, but they are no less important for that.

Other reasons for evaluation are somewhat less admirable. Webster and Stufflebeam's study of 35 very large school districts, for example, yielded one clear illustration of a principal asking for a survey which would make the district look good. The request appears to have been born of desperation since he was subsequently fired. We were told in an interview with one site that the administrator occasionally asks the evaluation unit to investigate some problem for which a decision has almost been made and the evaluation unit's task is merely to collect evidence to prove the case.

The interest in "puff pieces" or "hatchet jobs" and their frequency, however, appears to be rather low from our site visits and other information. It is, we believe, more likely in locally generated evaluations of personnel, for example, than in evaluation of federally supported programs. But evidence on any of this is weak.

We do not mean to imply that there are no problems in assuring integrity in the conduct of an evaluation, or in assuring candor in reporting. The matter is discussed in Chapter 5 on how well evaluations are conducted.

Other Ways of Characterizing the Functions of Evaluation

The questions enumerated earlier are a reasonable framework for understanding why an evaluation might be mounted. There are other reasonable classification schemes and other vernacular is often used to describe what evaluation is and what its purposes are. Specialization in evaluation has led to other terms to describe what an evaluation is supposed to accomplish. These are considered in the following remarks because they are frequently used by the evaluation community and occasionally in law and regulation.

Needs assessment surveys are descriptive studies undertaken prior to creation or modification of a program to establish what needs are, their scope, and their severity in particular target groups. The work may involve a few case studies or an opinion paper in the simplest case. The more elaborate may include formal statistical surveys as well as case studies. Illustrations from contracts completed under support of OE's Office of Evaluation and Dissemination during 1976-78 include "Assessment of Available Resources of Services to Severely Handicapped Children." Such surveys cannot normally be used to estimate the effects of programs in the least equivocal way possible.

Process evaluation include studies of the activities, operations, organization, and other aspects of a program. If the program is new, the examination is often labelled formative evaluation and implies troubleshooting activity. Illustrative projects supported by the OED include assessment of Emergency School Aid Act Program operations. Less elaborate varieties of this type of evaluation may involve short site visits and managerial case study. The more elaborate approaches can involve conscientious measurement of how many individuals receive services or how many institutions comply with instruction, of the degree and type of service or compliance. The main focus of such activity is understanding delivery of the program of services, and adherence to standards, rules, or instruction, rather than understanding or estimating effects of service. The idea of process evaluations is ambiguous in the sense that it can be regarded in most routine form as administrative monitoring through conventional record systems.

Outcome evaluations attempt to estimate in the least equivocal way possible the direct effects of a program on its main target group, usually children. Where there are multiple target groups and many effects are indirect, the activity may be labelled summative evaluation or impact evaluation. Where both costs and effects of program variations are estimated, the activity may involve cost benefit analysis. Illustrations of the type include NIE supported field experiments on career education programs and OED supported field experiments on some of the programs mounted under the Emergency School Assistance Act.

Administrative audits are often undertaken by state offices to verify that programs said to be operating at the school district level are indeed operating. In Florida and elsewhere, they are directed toward examining administration and processes and may involve visits to classrooms.

Administrative and technical support studies are often supported by evaluation budgets. These include conferences, technical assistance centers of the kind supported under Title I, special planning studies, development of new methodology.

Ambiguity in What Is Meant by Evaluation

The law rarely asks for specific types of information and never asks that a specific question be answered. Furthermore, written and readily available background information on the origins of the legislative demand, of the sort which appears in Hearings, for example, is often terse. Partly for these reasons, opinions often differed about what is intended, what evaluative information is essential, and how such information will be used. The differences have been registered in interviews with federal agency officials and in their professional papers, by at least some of the Congressional staff to whom we spoke.

Not recognizing the variation in interpretations of what is meant by evaluation is imprudent at best and dangerous at worst. The ambiguity is complicated by different levels of expertise and by different views of the topic within Congress and the agencies and across federal, state, and local government. Consider the following illustrations.

(a) The Director of Research in an education division of the federal government told us during an interview that his division did no evaluations. This was despite a list of projects for FY 1977-79 which included four items with the word evaluation in the title. He said they were "not really evaluations." His superior, a deputy assistant secretary, who was interviewed immediately afterward, said that "almost every project we do is an evaluation."

(b) In an interview with a Congressional staffer, we were told that his Committee was interested in effectiveness, that is, how many children are in the federal program in primary school, junior high and high school. In response to the question "what about effects on children" he said that this is not what is normally meant, in his judgement, by effectiveness among his Committee members even if it is viewed that way among staff and members of other committees. This discussion is consistent with difficulties encountered in the NIE Compensatory Education Study and the long negotiation needed to settle on the major questions to be addressed. It is consistent with the confusion and "protracted negotiation between USOE and concerned Congressional committees" over models for evaluating Title I programs.

(c) In a presentation before a National Academy of Sciences Committee, a Congressional staff member criticized the studies supported by a federal educational agency as being "silly evaluations." It was pointed out that the studies were more properly characterized as basic research, silly or not, and the staff member responded by saying that he did not distinguish between research and evaluation.

(d) In a recent study by Syracuse Research Corporation of Title I supported programs, the majority of respondents in a large sample said they regard evaluation as "monitoring and testing students." Estimating the effects of Title I programs, a function implied by early government requirements for evaluation, was apparently of minor interest. The perception of evaluation as "testing students," or in special education programs as diagnosis of children, or in bilingual education as evaluating materials, was confirmed in our round-table discussions and some site visits.

(e) The bewildering array of senses in which the phrase "program evaluation" is used in the professional literature forms the stuff of a 17 page article by Gene Glass of The University of Colorado and Frederick Ellett of UCLA. They plough through, with alarming conscientiousness, evaluation as "applied sciences," "systems management," "decision theory," "assessment of progress toward goals," "jurisprudence," "description or portrayal," and "rational empiricism."

(f) At least a few federal agency staffers reckon that Congress passes the buck to the federal agencies to instruct the Congress about evaluation at least in some instances. For example, one of our informants at NIE asked a Congressman what he wanted from the program and its evaluation. He left it up to the evaluator, saying, "We just want to goose the system." It's for reasons of this sort that one CBO staffer rarely uses the word evaluation and prefers to make explicit the particular questions which are being addressed in an evaluation.

(g) Ambiguity is not confined to education. The most recent UNESCO conference on evaluation techniques in Asia confined virtually all attention to organization of computer systems rather than to evaluation in the sense of determining whether programs operate as they should or in the sense of estimating impact. About a half dozen different definitions of evaluation seem to have been used in recent Pan-American conferences on evaluating nutritional programs.

There are a number of reasons for ambiguity in popular and professional use of the word evaluation. A fundamental one is simply that the word implies different things to different people. The differences stem partly from the different disciplines involved. In education, for example, it is not unusual to encounter management, sociology, statistics, psychology, accounting, and policy sciences represented in an evaluation. Differences also stem partly from the variety of methods used by a discipline to examine a problem. The administrator may bring case studies to bear. The statistician normally argues for statistical evidence. The lexical inventiveness of academic scholars, bureaucrats, and politicians complicates matters further. Despite the utility of some labels and slogans, the introduction of "self-referent evaluation," "illuminative evaluation," "responsive evaluation" and other neologisms is confusing. In some instances, the label attached to the activity hinges on people, regardless of title: If A does it, it's evaluation; and if B does it, it's research.

Finally, it occasionally appears to be prudent to keep the word vague. The demand to "evaluate" is sufficiently general to permit investigation and audit without inciting the fears that these activities do. It has been used as a synonym for research when the word research is found offensive by legislators or parents.

None of this is peculiar to educational evaluation judging from recent work by William Kruskal at Chicago and Frederick Mosteller at Harvard. Terminology in statistics is used popularly to give scientific dignity to simple information collection. For example, a writer may announce that he has taken a "random sample" to lend scientific legitimacy to his efforts and we find later that the word haphazard is rather more accurate. Uses of words like "experiment," "audit," "psychotic," and so on are used promiscuously despite professional communities that ascribe to more or less explicit definitions in each case.

We stress the matter here because the ambiguity does affect communication at federal, state, and local levels of government. It affects quality of evaluation, costs, and so on. Some of our recommendations bear on the problem.

2.2 AUDIENCES FOR EVALUATION RESULTS

In any potential audience for evaluation results, there are pockets of sturdy indifference as well as pockets of remarkable interest. The following remarks illustrate this variety at national, state, and local levels of governance.

National Level

The GAO's recent work classifies the audiences for evaluation into three groups: policy-makers, managers, and oversight agencies. Various public interest groups which occasionally demonstrate an interest in evaluation, advisory groups, program staff, and parents, and a sizeable community of evaluators constitute two nongovernment audiences. Not all members of these audiences are equally attentive.

In principle, the relevant policy-makers include the Congress, since the demand for evaluation of ongoing programs and many new ones is made in law. The case studies on use of evaluation in Chapter 6 suggest that Congress is indeed an audience at least at times. There are explicit references to evaluations in committee reports, in the incorporation of evaluation findings into bills, and in the rationale provided for appropriations. The pertinent cases which include the clearest evidence are the NIE Compensatory Education Study, the National Day Care Study, Rand's Study of Federal Programs Supporting Educational Change, Bilingual Education, the Fund for the Improvement of Post-secondary Education, and of Title I Testing. Evaluations of Follow Through among others are remarkable for the lack of audience reaction, but notable for their use in provoking discussion.

The reports of the Senate Appropriations Committee, the House Education and Labor Committee and others register the interest of this audience but indicators of the interest are not always uniformly clear. For instance, the Senate Committee's 1979 Report enumerates appropriations and gives a brief rationale for each appropriation. Of the 85 education budget items the Committee considered, 11 include reference to evaluation as part of the rationale for support, though evaluations are available and were used we believe in at least 6 more cases. The absence of evaluative data is mentioned twice, though probably over 20 items in the catalog have had no formal evaluation at all.

Our sources of information about Congressional staff interests include a recent survey by Florio, Behrmann, and Goltz of 26 staffers of subcommittees and committees dealing with education. These include the House Committee on Education and Labor and five pertinent subcommittees, the Senate Committee on Labor and Human Resources and four of its subcommittees, Senate Appropriations and Governmental Affairs Committees. Of the 26 staffers identified as an audience in principle, seven considered the influence of evaluation on their work to be significant, eight said importance depended on the issue, and eleven said evaluation was not useful to them in the form delivered. We conducted interviews with eight of these staffers, four nominated by Congresswoman Holtzman's staff, the remainder by ex-staffers, in the interest of independent review. All were informed about evaluation in some degree. But their own opinions about how useful evaluations are hinged considerably on particular evaluations--some being regarded as useful, others not--and their definition of "useful." One staff member, for instance, said that evaluations were not useful to his committee but also said clearly that evaluations were used to guide Committee members' questions during Hearings. Their remarks on how evaluation results are used have been incorporated into Chapter 6.

The most relevant Congressional support staff include four members of the Congressional Budget Office who capitalize on educational evaluation as a policy tool. The persons to whom we spoke regard themselves as discriminating consumers of evaluations, and use evaluations in policy development. The case study on CBO in Chapter 6 illustrates their use of results. Members of the Congressional Research Service with responsibility in education in principle constitute an audience for evaluation. The two people to whom we spoke regarded themselves as brokers of evaluation and research.

In principle, federal agency program managers constitute an audience for evaluation. This involves a presumption that the manager will act upon the findings of an evaluation supported by the division of evaluation, by GAO, or some other reasonably independent agent. In practice, there is confusion: Evaluation information requested by Congress is not necessarily useful to the manager and evaluative information needed by the manager is not necessarily useful to Congress. Advertising that "evaluation is a management tool" as OMB has done implies, to at least some policy analysts, that it will then be less useful to policy-makers.

In practice, there is also competition between the program unit and the evaluation unit for funds and Congress's attention. Some of the anomolous possible consequences of competition here can be illustrated by a Congressional staffer's opinion that federal managers do not take action on evaluations and as a consequence the evaluation unit's budget, not the program budget, is reduced. Under the new Department of Education, the evaluation unit is lodged in the Office of the Assistant Secretary for Management. That Office will, its executives expect, help to link evaluation more closely to program management without degrading the extent to which evaluation can meet Congressional requests.

The case studies on uses of evaluation given in Chapter 6 reflect some management interests in the information. The verifiable uses include changes in federal regulations produced by the National Day Care Study, the NIE Compensatory Education Study, smaller scale assessments of Title I testing programs, and the Rand evaluations of Federal Programs Supporting Educational Change. The unwillingness or inability to use some evaluations for legitimate and other reasons is reflected in the Follow Through case study. The most consistent but small agency audience for evaluations which estimate effects of programs is the Joint Dissemination Review Panel, a 22 member board that included OE and NIE staff. The JDRP routinely reviews programs and evidence submitted by program developers at the local, state, or federal level to determine if evidence is sufficient to warrant providing the program with an opportunity to apply for dissemination grants.

The priority accorded to evaluation by federal managers varies. In bilingual education, for example, the reviews of grant proposals involve a 110 point scoring scheme; fifteen points go to evaluation, suggesting that it cannot play a decisive role in funding decisions in the division of bilingual education. Judging from public speeches by some bilingual education experts, the recent AIR evaluation of bilingual education has provoked more interest in evaluation for the sake of self-protection if nothing else.

In Vocational Education, the law requires the state board responsible for program administration to provide an accountability report to the Commissioner annually. The specification of content practically legislates management as an audience, in that it must contain a "summary of the evaluations of programs...and a description of how the information from these evaluations has been or is being used" (Section 108, Title I - Vocational Education). We are unaware of any formal assessment of these reports or any substantial way in which they meet the requirement.

At the GAO, there is a notable interest in evaluation among members of the Program Analysis Division, a unit was created in response to law requiring that GAO play a major role in overseeing evaluations. The interest and some of the expertise is gradually finding its way into other divisions of the agency. The approaches being developed by GAO to guide evaluations are based at least partly on earlier evaluative research supported by the National Institute of Education and National

Science Foundation. In this sense, GAO is an audience for evaluative methods. GAO is also an audience, at times, for the product of evaluation itself if we may judge by GAO's use of evaluation by other agencies. A case in point, the National Day Care Study, is discussed in detail in the chapter on uses of evaluation results.

The courts are not a regular audience for evaluations. But educational policy research, supported at least partly by the federal government and directed at questions which evaluations address, is used episodically. Information generated by an OED-supported evaluation of bilingual education programs has been introduced as evidence in a New York state court case, Cintron vs. Brentwood School District on access to bilingual education. In the Supreme Court, Bakke vs. University of California employed data from the OED-NCES supported National Longitudinal Study of the High School Class of 1972. Applied research on the effects of integrated classrooms on children's achievement, conducted by university researchers in pursuit of scholarly interests, has been admitted as evidence in Swann v. Charlotte-Mecklenberg Board of Education, in Hobson v. Hansen (Washington, D.C.) and Keyes v. School District No. 1 (Denver). At the state level, investigations by Coleman suggest that in California research on the effects of desegregation continue to be used partly as a result of a State Supreme Court ruling on admissibility of the evidence. Methodological research on evaluation, produced with federal support, is also of occasional, direct interest. For example, the Federal Judicial Center has created a Committee on Experiments in the Justice System to understand how sophisticated evaluation methods, notably randomized field experiments, can be used to determine whether and how well innovations work. The Committee, chaired by Judge Edward Re, has included meetings at which methodological work supported by the National Institute of Education as well as National Science Foundation has been formally presented. We have not had the resources for a detailed examination of how often information generated in the national level evaluations has been used by the federal, state, or local courts. The preceding remarks merely illustrate the beginnings of interest in the judicial sector.

The Local and State Interest in National Evaluations

Our site visits to local education agencies generally revealed little awareness about evaluations produced at the national level. Interest generally focuses on locally executed evaluation if there is any interest at all. But there are two important exceptions.

The first exception involves large school districts which often have research and evaluation units and states with well developed evaluation practices. Those units are consumers of some evaluations and of methodological research on evaluation which finds its way into the professional journals. So, for instance, the input evaluation unit of the Dallas Independent School District's research and evaluation division is responsible for monitoring national reports. The examination of reports of Title I evaluations led eventually to Dallas's field testing of an instructional approach identified as successful, DISTAR, in the evaluation report.

A similar exception holds for states with well developed approaches to evaluation of state and federally supported programs. The usefulness of this information to state staff is partly rhetorical—to persuade someone of the value of the program--and partly managerial--identifying good and bad practice. For instance, Iowa's Oliver Himley cited three national evaluations bearing on Title I programs in 1979 Oversight Hearings of the Subcommittee on Elementary, Secondary, and Vocational Education. He appeared to be the only one of representatives from nine states who knew about the studies. Kansas and California have both used the OED-supported Rand Study of Federal Programs Supporting Educational Change to modify legislation (see the case studies in Chapter 6). And in other states, such as Michigan and Minnesota, we did encounter program staff and evaluators who were aware of national evaluation.

A second major class of exceptions to local disinterest or lack of awareness about national evaluations concerns evaluations which receive considerable professional or popular press coverage. At Site J, for example, the director of bilingual education was familiar with the recent national evaluations of the bilingual program undertaken by OED. The professional debate over the national evaluation itself led this director to be concerned about the quality of the report that she received from the contractor she had hired to evaluate Site J's program, and she asked us for comments. The national evaluation has been covered by professional periodicals such as Phi Delta Kappan and by network television. NBC's Tom Snyder (May 9, 1978) included interviews with federal program managers and Ohio Representative John Ashbrook. The Rand Corporation's evaluation of "Federal Programs Supporting Educational Change" is another case in point. The report itself was covered by at least a half dozen local newspapers and at least one syndicated columnist. Press coverage did provoke discussion, though it is not clear how it may have influenced decisions. Spencer Rich of the Washington Post covered an NIE supported evaluation of Jesse Jackson's PUSH-EXCEL program (April 22, 1980). The article, a balanced one, evoked considerable interest in both the program and evaluation judging from letters to NIE.

Documenting instances in which popular treatment of professional evaluation has provoked action is time consuming, and the resources available to this project have not been sufficient to do so. No formal newspaper clipping system is employed in any federal educational evaluation unit we've visited, though there is informal attention to this topic.

Local and State Audiences for Local Evaluations

The most recent report of a large scale survey bearing on audiences for evaluation was issued in 1979 by UCLA's Center for the Study of Evaluation. Their effort focused on large school districts with research and evaluation units. Respondents in that survey, directors of research and evaluation units, reported that the most consistent users of their reports were superintendents and central office staff of the school district (60%) and principals (53%). About 30% of the directors said that teachers were consistent users of the information they generated. School board members are still less likely to be a regular user of the information and parents are reported to be least likely to be a consistent user group (9%). The UCLA survey covers all evaluations generated by a research unit, not only the federally supported ones.

From our own site visits, we conclude that school board interest in local evaluation varies considerably from school district to school district and state to state. The extremely active audience is exemplified by the school district in Site E, for example, where we were told that one school board member, who on receiving any proposal for funding immediately flips to the budget and the line item for evaluation, and who on receiving reports about a program asks where the evaluation is, and who calls district evaluators to obtain information about evaluations directly and to remind evaluators that they've missed a deadline. The school board itself at Site J has required evaluation of some activities every three years, executive summaries of evaluation in plain English. In Dallas, the school board has a formal program evaluation committee which includes board members to whom research and evaluation reports are regularly provided. Such interest among school boards is, we believe, exceptional. For instance, we found no evidence that Title I evaluations in Site A are considered seriously by the school board, though the attention given evaluation by the Title I director is clear. Site J's board involvement is not significant, judging from our interviews, because management and budget issues are higher in priority.

For Title I programs, at least one local audience is implied by federal law. Parent Advisory Committees are given responsibility at the local level for providing advice on implementation of Title I programs. As a matter of practice, the audience sometimes has less access than it should to evaluation reports and it may be uninterested. The case study given in Chapter 6, on use of evaluation findings by Parent Advisory groups, covers both active and inactive segments of this audience.

In any given site, the audiences range considerably depending on the nature of the information produced by the evaluator, and the relations between evaluator and various groups within the district. Title I evaluation in Providence, for instance, suggests the following pattern of information needs being satisfied. Systematic assessments of the relation between instruction time and achievement was eventually used by Title I teachers and administrators. Assessments of clerical errors in records, of cut off points on tests used to assign children to different instructional regimens, were used by program management. Assessments of class size, of the appropriateness of tests, of classification problems engendered by tests, of the appropriateness of program objectives based on tests are provided to teachers and used by them. The Title I Parent Advisory Committee appears to be a consistent audience for evaluations showing poor performance in middle schools relative to elementary schools and has actively used the information in pushing for program modification.

In other sites, such as Site J and Site F, the federal emphasis on achievement levels of students is often of considerably less interest to program managers, superintendents, and school boards, than is information about children's attitudes, children's self-concepts, and parental reactions to programs.

Consumer and Producer Education

One clear implication of this and others' investigations is that the audience for evaluation shifts considerably in membership and the need to inform the audience about purposes of evaluation is a recurring one. Site J, for instance, has had five superintendents in ten years; the veteran school board member has had six years experience and majority have less than two. Site A's school board has had over the past five years a dramatic shift in membership and part of the program evaluator's problem in that city is to overcome both ingenuous and informed suspicion of the new board against district staff. At the state level, changes in membership of the legislature and legislative staff can be rapid. This puts substantial demands on state evaluation staff in California, for instance, to explain the purposes, origin, nature and conduct of evaluative activities. At the national level, Congressional staff member turnover is relatively high. Though there remains a core of individuals who are sophisticated about the flaws and benefits of evaluation, the audience changes often enough to put considerable demand on communication efforts. The problem has been similar with the Education Division of DHEW. Eight Commissioners of Education in a ten year period and a small army of migratory executives make matters difficult. Briefings on evaluation by the OE's Office of Evaluation and Dissemination have become routine and elaborate in response to the problem.

The audience is not the only group being educated. The difficulty of communicating about evaluations to a heterogeneous, often nontechnical audience, has had the benefit of spawning new interest in ways of presenting information. Evaluation staff of the Austin Independent School District, the Baltimore County Public Schools, the Philadelphia School District have been remarkable in this respect. Austin, for example, has managed to produce a thoughtful manual on reporting, based on hard experience. The experience in smaller districts and in districts where evaluation staff are unable or unwilling to provide information about the way they approach the problem is almost invisible. In some cases, it is not especially good. Parent Advisory Councils in some areas have notable difficulty in accessing reports, for example, and in understanding what is being said on account of technical language. It is partly on account of such difficulty that the UCLA Center for the Study of Evaluation has created a newsletter to periodically report on effective strategies for getting information used at the local level. The production of the periodical, U.S.E. (Using School Evaluations), is supported by the National Institute of Education.

2.3 LEGISLATION ON EVALUATION

A fundamental reason why evaluations are done is that Congress wants them done, if we may judge from law and from our interviews at local, state, and federal levels. To understand the scope of legislative demands for evaluation, we undertook a search for statutes bearing on the topic and a brief analysis of them.

The main vehicle for search was the LEXIS computerized legal document retrieval system, using Title 20 of the United States Code. Scope of search was restricted to those educational programs most relevant to this Project, e.g., elementary and secondary education, vocational education, education for the handicapped, and so on. Graduate or higher education programs, school nutrition programs, and library programs, were excluded. Moreover, the search focused on statutes containing the word evaluation. It then excludes evaluative activity which may have been labelled differently. Details of the search, problems encountered, and detailed analysis are given in the Appendix to this report. We believe it is the only attempt since the GAO's in 1974 to map this terrain.

Recency of Amendments and Enactments

All the subsections of the U.S. Code containing references to evaluation have been placed in the Code or amended since 1968. Furthermore, 73% of the subsections were amended by the 95th Congress in 1977-78. Over 60% of the subsections were amended or enacted by a single act of legislation, the Education Amendments of 1978 (Public Law 95-561). The number of citations to evaluation subsections of the U.S. Code by year are as follows:

Year	1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78
Number	1	0	5	0	7	0	7	5	8	11	75

Character of General Provisions

The General Provisions Concerning Education apply to all federally funded education programs "for which an administrative head of an education agency has administrative responsibility" (§1221(b)). They further specify that any state or local application for federal funds for an education program must contain an evaluation component and assurances of cooperation in providing data for federal level evaluation efforts. The local agency must report to the state and the state to the Commissioner of Education (§1232d; 1232e). The general purpose of such evaluations is to "determine the effectiveness of covered programs in meeting their statutory objectives (§1232d).

At the federal level, there is overlapping and potentially conflicting authority for evaluation of education programs. The Secretary of DHEW has had a similar statutory mandate with no clear indication how these evaluation efforts are to differ from those of the Commissioner of Education (§1226c). Further, the Comptroller General may conduct special evaluations of programs upon request by Congress (§1227). The National Institute of Education and the National Advisory Council on Indian Education also may evaluate educational programs within the province of their concerns (§§1221e; 1221g).

These statutory sections represent the most general evaluation requirements and they are stated in rather broad terms. Various statutory sections require needs assessments (§1226c), outcome evaluations (§1226c; 1231a), and even cost/benefit analyses (§1226c). Another section of the General Provisions requires the Commissioner of Education to conduct a comprehensive study of evaluation practices and procedures at the national, state, and local levels for federally funded elementary and secondary education programs and this Project stems from the requirement (§1231a). While these sections of the General Provisions establish the cornerstone of federal policy on educational evaluations, the influence of these requirements remains unclear. The General Provisions are to apply unless otherwise specified by federal statute, and no such specification was encountered in the course of this research. Further information and citation to specific subsections of the General Provisions are provided in the Appendix to this report.

Specific Programs

To understand evaluation requirements for specific programs consider statutes on education of handicapped children, bilingual education, Title I, and vocational education.

The Education of Handicapped Children (§§1401-1461). Since 1975, the Commissioner of Education has had primary responsibility to conduct evaluations of special education programs. Either directly or by grants or contracts, the Commissioner has been required to assess the need for special education programs, the extent to which needed services are delivered, and the effectiveness of these programs in meeting the goals of special education (§1418). State and local education agencies must comply with these regulations as a condition of federal funding of special education programs (§1412). Each year, the Commissioner is to file a report with Congress on the progress being made in providing special education services. This report is to include a detailed description of all evaluation activities conducted by the Commissioner and an assessment of the effectiveness of each education agency in providing special education services in the least restrictive environment and in preventing erroneous classifications (§1418). The Commissioner must also evaluate the effectiveness of each of the model centers and experimental programs in meeting the special needs of handicapped students (§1425).

Bilingual Education (§§3221 - 3261). The statutory mandate concerning evaluation of bilingual education programs is vague and seems to reflect a cautious approach in establishing rules for a new education program. In fact, in the statutory statement of policy Congress notes that research and evaluation capabilities in bilingual education need to be strengthened (§3222). The Commissioner of Education has been directed to develop models of evaluation to be used in assessing the progress made by participants in attaining English language skills. The local education agencies then implement the evaluation models and report to the Commissioner as a condition of federal funding (§3221). The Commissioner reports to Congress once every two years concerning the need for bilingual education and the success of the federal programs in meeting this need (§3241). While these initial evaluation efforts are underway, Congress requests the Secretary to develop more sophisticated evaluation and data collection models by September 1980 (§3241). The National Institute of Education is also required to develop and evaluate effective models for bilingual education (§3252). Further information and citation to specific subsections may be found in the Appendix.

Evaluation of Title I Programs (§§2701 - 2854). The statutory standards for evaluation of Title I programs delegate primary responsibility for evaluation of Title I programs to the local agencies (§2734). The state education agency has been directed to provide technical assistance to the local agencies, and to compile the findings of the local evaluations in a report for the Commissioner of Education (§2822). The Commissioner, who must also provide technical assistance to local agencies, then must combine the findings of the state reports with national evaluations of Title I programs and present a biannual report to Congress (§2833). The National Advisory Council on Quality in Education also may evaluate Title I programs (§3171). Further information and citation to specific subsections are given in the Appendix.

Vocational Education (§§2301 - 2461). The core evaluation requirements for vocational education involve the common scheme of the state education agency reporting to the Commissioner of Education, who reports to Congress. At the state level, authority to conduct evaluations is retained by the State Advisory Council, which in principle must meet detailed standards in evaluating local vocational education programs (§2305).

At the federal level, the authority to conduct evaluations of vocational education programs is divided among several agencies with overlapping and potentially conflicting mandates. In addition to evaluation of vocational education programs by the Commissioner of Education, federal level evaluations are to be conducted by at least three other groups. The National Advisory Council on Vocational Education is authorized to conduct independent evaluations of state vocational education programs and to file an annual report with the Commissioner of Education, the Secretary of Labor, the Congress and the President (§2392). The National Institute of Education has authority to "study and evaluate" a broad range of vocational educational programs in order to make recommendations for the redirection and improvement of vocational education programs in the next decade. NIE also has the authority to conduct "not more than three experimental studies" to achieve the purpose of the evaluation (§2563). Finally, the Commissioner of Education and the

Secretary of Labor have joint authority to evaluate bilingual vocational training programs and file an annual report with the President and Congress (§2412).

In addition to these primary evaluation mandates, a number of secondary mandates for research in vocational education are included. A national center for research in vocational education was established to develop "methods of evaluating programs, including follow-up studies of program completers and leavers" (§2401). A Coordinating Committee on Research in Vocational Education is established to develop "an effective management information system. . . to achieve the best possible monitoring and evaluation of vocational education projects" (§2304). Finally, the Commissioner of Education is to study sex discrimination and sex stereotyping in vocational education programs (§2563).

Other Strategies of Evaluation

In addition to the General Provisions and the four major education programs discussed above, evaluation standards for twenty-six other education programs were examined. The standards vary widely across programs. For some education programs, the local agency is responsible for developing appropriate evaluation models to be included in the application for federal funds (e.g., Metric Education -- §2953; Community Schools Program -- §3288; Dropout Prevention Programs -- §3387). Other education programs place heavy reliance on contractors or grantees to conduct the required evaluations in accordance with standards developed by the Commissioner (e.g., Media Education -- §541; Consumer Education -- §2983). Some education programs offer no guidance beyond the general admonition that the federally funded program should be evaluated (e.g., Educational Improvement and Resources Support -- §3084; Gifted and Talented Children Program -- §3315).

Types of Evaluation Activities Required by Education Statutes

To obtain better understanding of the legislative references to "evaluation," the references were classified as indicating one or more of the following: needs assessment; process or formative evaluation; outcome or summative evaluation; cost/benefit analysis; and an "unspecified" classification. "Needs assessment" was defined as those activities directed toward determining the nature or extent of a problem, such as a survey to determine the need for a bilingual education program among migratory children. "Process or formative evaluation" was defined to include an examination of the nature of the services being provided or an examination of the functioning or operations of the education program. "Outcome or summative evaluation" was defined as those activities designed to determine the effect of a program on the problem it was intended to solve, or an assessment of the effectiveness of a program in meeting the purpose of the statute. "Cost/benefit analysis" was defined as those activities which compare estimates of the impact of a program with the costs of providing the services. Any evaluation requiring a comparison between the funds spent on a program and the results of the program was classified as a "cost/benefit analysis." The "unspecified" category included those references to evaluation in which the nature of the required activities could not be determined from the context of the statute. Finally, an "other" category was included to permit examination of evaluation mandates which did not fit into any of the anticipated classifications.

Explicit statements concerning evaluation of the impact of an education program are much less frequent than more general statements concerning evaluation of the effectiveness of the statute. Only eight of the twenty-nine programs or levels of programs for which the statute implies outcome evaluation either mention or imply an evaluation of the impact of the education program on the perceived problem (Education of Handicapped Children, State and Local Evaluation -- §1413(a)(7); Federal Evaluation -- §1418(a); Bilingual Education, Local Evaluation -- §3241(d); Vocational Education, State Evaluation -- §2312(b)(1)(B), Federal Evaluation of Bilingual Vocational Programs -- §2412(a)(2); Preschool Partnership Program -- §2917(b)(3); Biomedical Sciences Program -- §3054(a)(11); Emergency School Aid -- §3200(a)(11)). The remaining twenty-one programs or levels of programs requiring outcome evaluations contained a statement concerning an "evaluation statute." This phrase occurs with such consistency that it has the character of statutory boilerplate, employed when there appears to be general interest in the functioning of the program but there is no intention, consensus, or clear preference for "outcome or summative" evaluations rather than "process or formative" evaluations. If one assumes that the use of this general phrase concerning evaluation of the effectiveness of the education program only expresses an interest in requiring some form of objective assessment of program operations, then these more general mandates can be taken out of the "outcome or summative evaluation" category and combined with the "unknown" or unspecifiable evaluation mandates. Only eight of the forty programs which require an impact assessment can be classified as requiring "outcome evaluations," and thirty of the forty programs contain one or more of the general mandates which does not specify an evaluation question.

If general statements about effectiveness in achieving purposes of the statute are classified as outcome or summative evaluation, then this is the most common type of mandated evaluation activity. In each major education program and in seventeen of the twenty-six minor education programs, the purpose of the evaluation requirement was either to determine the impact of the program or to determine the effectiveness of the program in achieving the purpose of the statute. The nine minor programs which did not require outcome evaluations all had "unspecified" evaluation requirements, suggesting that the absence of language requiring an outcome evaluation did not imply a preference for one of the other evaluation models.

In six of the forty programs or levels of programs, the evaluation mandate referred to a needs assessment of some kind. Typical of these mandates are those for special education in which the Commissioner of Education is required to determine the number of children in each state with particular kinds of educational disabilities (Education of Handicapped Children, Federal Evluation -- §1418(b)(1)). Only some of the statutory mandates for needs assessments are referred to as evaluation. In six different programs or levels of programs, similar data collection mandates were required without using the term "evaluation." All four major education programs and the General Provisions require needs assessment of some form by some level of government. This suggests that needs assessment is recognized as a valuable data collection activity, though it may or may not be referred to as "evaluation."

Five of the forty programs or levels of programs required a "process evaluation." The statutory language mandating a process evaluation varies greatly. Examples include the mandate to special education programs to evaluate the effectiveness of procedures intended "to assure that handicapped children receive special education and related services in the least restrictive environment . . ." (Education of Handicapped Children -- §1418 (d)(2)(A)), and the mandate to the National Institute of Education to "conduct an evaluation and study . . .[to analyze] the means of assessing program quality and effectiveness" (Vocational Education, Federal Evaluation by National Institute of Education -- §2563(b)(1)(C), also classified as a mandate for an outcome evaluation). Four of the five mandates for process evaluation applied to the major education programs.

Three of the forty programs or levels of programs discussed evaluation in terms of a "cost/benefit analysis." Two of the four major education programs and the General Provisions require some form of cost/benefit analysis at some level of the program. Only programs in Bilingual Education and Education of Handicapped Children have no such requirement. Typical of such evaluation mandates is the requirement that state evaluations of Title I programs determine the "effectiveness of payments in improving educational attainment" (Title I Programs, State Evaluations -- §2822). Two instances were found which expressed similar mandates but which did not use the term "evaluation" (Vocational Education, State Evaluations -- §2308(b)(2)(B); Career Education Incentive Program, State Evaluation -- §2613(b)).

In fifteen of the forty programs or levels of programs, the nature of the intended activities could not be determined from the context of the statute. This lack of specificity was more common in the minor education programs and typically took the form of a general statement such as, "All projects shall include an evaluation component" (Correction Education -- §3032(a)). Occasionally the evaluation requirement was simply listed along with a number of other required activities, such as "research and evaluation" (Educational Proficiency Standards -- §4443(a)(3); Law-related education -- §3002(d)(5)).

Four of the forty programs or levels of programs discussed evaluation in terms which did not fit any of the above categories. In two instances, "evaluation was used to indicate the need for diagnostic testing or individual assessment (Education of Handicapped Children, State and Local Evaluations -- §§1412(2)(C), 1414(a)(1)(A), 1415(b)(1)(A); Basic Skills Improvement Program, State Program -- §2902(d)(6)). In one instance the term "evaluation" was used to describe the assessment of consequences for education programs of changing the statutory definition of the word "Indian" (General Provisions for Educational Programs, Federal Evaluations -- §1221h(b)(3)).

Types of Evaluation in Different Levels of Government

Just as the term "evaluation" can have different meanings across different education programs, "evaluation" can have different meanings both within and across different levels of government concerned with administering a single program. For example, special education programs required four different kinds of data collection, all described as "evaluation." For the state and local level, the word "evaluation" is used to indicate the need for outcome evaluations and diagnostic testing of individual students (Education of Handicapped Children, State Evaluations -- §§1413(a)(7); 1412(2)(C); 1414(a)(1)(A); 1415(b)(1)(A). For the federal level, the word "evaluation" implies needs assessments and process evaluations, as well as outcome evaluations (Education of Handicapped Children, Federal Evaluation -- §§1418, 1425). The federal level requirements in the General Provisions use the word "evaluation" to indicate needs assessments, outcome evaluations, cost/benefit analyses, one "unspecified" evaluation, and an assessment of the consequences of changing the legal definition of the term "Indian." (General Provisions for Education Programs, Federal Evaluations -- §§1226c(a); 1231a(a)(3); 1221h(b); 1221(b)).

Despite the variety of meanings of the term "evaluation" in the mandates for data collection by single level of government, some conclusions can be drawn regarding the kinds of evaluation activities which are commonly assinged to specific levels of government. Examining only the requirements of the General Provisions and the four major educational programs, one finds the responsiblity for conducting needs assessments and process evaluations most commonly located at the federal level (Education of Handicapped Children, Federal Evaluations -- §1418(b)(1), 1418(d)(2); Bilingual Education, Federal Evaluations -- §3241(c); Vocational Education, Federal Evaluation by the National Institute of Education -- §2563(b)(1); General Provisions for Educational Programs, Federal Evaluations -- §1226c(a)(2)). When responsibility for conducting needs assessments and process evaluations were located at lower levels of government, these duties were not defined as evaluation (Title I Programs, Local Evaluations -- §2734(b); Vocational Education, State Evaluations -- §2551(b)). Mandates for outcome evaluations and cost/benefit analyses were found at both the federal and state levels.

2.4 FUNCTIONAL CHARACTER OF EVALUATION: STATISTICAL DESCRIPTION

National Level

One vehicle for understanding what questions have been addressed in federal evaluations is to focus on activities of OE's Office of Evaluation and Dissemination. The Annual Evaluation Report, issued by OED, carries descriptions of studies completed during the fiscal year. Reports for 1977, 1978, and 1979 were reviewed. The completed evaluation studies that were highlighted in each report were classified according to the questions that the studies addressed. The results are as follows:

Summary of Evaluations
Completed by OED in Fiscal Years 1977-79

Completed in:	1977	1978	1979
Total Number:	20	18	26
Percentage focusing primarily on:			
Who is served	60%	44%	46%
Nature and Cost of Services	95%	80%	91%
Effect on recipients of instructional services	25%	44%	35%
Costs or benefits of alternatives	6%	5%	0

The percentage of studies emphasizing each topic do not add to 100 because many studies have multiple purposes. Detailed exhibits are given in the Appendix.

The main inferences we draw from the table are that:

(a) Contrary to common views of the Office of Evaluation and Dissemination, most studies have not been directed at estimating the effects of programs on their major target groups. (See Chapter 3 on costs).

(b) The proportion of evaluations with a strong emphasis on examining costs or benefits of alternatives is puny.

The National Institute of Education's primary mission is research rather than evaluation of ongoing programs. The development work on new programs engenders activities which could be labelled evaluation, however. And the support of work on methods of measuring achievement on designing experiments, and on other topics are pertinent to evaluations undertaken by other federal agencies including the operating components of the new department. The 35 contracts and grants awarded by NIE and current in 1977, 1978, and 1979 were classified to determine how they relate to questions normally addressed in evaluations. The results are as follows:

Total Number of Contracts and Grants	35
Percentage focusing primarily on:	
Characteristics of children, intellective growth, etc.	20%
Nature of projects or programs	23%
Effects of new projects, program components, or program variations	49%

Detailed information is given in the Appendix. Again, most research projects have several main themes and so the percentages do not sum to 100. The major exceptions to the regular grant or contract research bearing on evaluation are the specially mandated studies such as the NIE Compensatory Education Study, the Safe School Study, Project Propinquity, and the Vocational Educational Study that is currently underway. These have been excluded from the count.

The Office of the Assistant Secretary for Planning and Evaluation in Education has, despite its title, invested most resources in planning and feasibility studies and in policy analyes rather than field evaluations. For the reports issued between 1976 and 1979, supported by contract, we found the following:

Total number of reports:	70
Percentage emphasizing:	
Descriptive policy analyses	41%
Planning and feasibility studies	24%
Data acquisition and processing	20%
Evaluation	7%
Technical Assistance	7%

Roughly speaking then, the agency most likely to investigate implementation has been the Office of Evaluation and Dissemination. Addressing questions about effects of programs is less frequent than addressing questions about who is served and how services are delivered. Questions about the effects of programs are more likely to be addressed by NIE, but the programs examined are generally new or experimental and small. ASPE produced few reports bearing on evaluation. These reports were produced through grants and contracts.

More generally, if we examined federal law for 29 major programs, we find in most of these, the language demands evaluation of the effectiveness of a program in "meeting the objectives of the statute." In the absence of other information we infer that any or all of the questions about who is served, the nature of service, and so on, could properly be addressed in an agency's efforts to implement the requirement. In a minority of statutes, there is a clear stress on assessing impact of the program on a particular problem. The provisions in Public Law 94-142, for instance, are more specific in asking that information be obtained on who is served, the level of need for service and the nature of services. The statute by itself does not assist us in understanding whether one ought to try to estimate the effect of those services on children. In six of 40 major and minor programs, there is an explicit demand for needs assessment and in three there is an explicit reference to cost/benefit analyses. For most programs, then, the language is general and cannot be used to judge which questions can be or should be addressed in a field evaluation. The statutes containing more specific language are exceptional, and usually concern special studies which the Congress wants undertaken.

Special Studies

The specially mandated studies are too few to make statistical description useful. Their aims vary considerably but address one or more of the questions outlined earlier. Consider the following illustrations.

The Safe School Study, for example, was undertaken by NIE at Congress's direction to "determine the number of schools affected by crime or violence, the type and seriousness of the crimes, and how crime could be prevented." It was mandated as part of the Educational Amendments of 1974 (Public Law 93-380). The report, completed in 1978, is based heavily on surveys. Its specific origins lie partly in initiatives by Representative Bingham of New York and Bell of California and by Senator Cranston of California.

The NIE Compensatory Education Study was, required, under the law, to examine purposes and operation of the Title I program and to analyze its effectiveness. The focus on operations, including alternative allocation formulae was substantial. The origins of the study lie partly in general concerns about the absence of good information on Title I performance and special interests of Representative Quie's in alternative approaches to allocation, judging from the hearings preceding enactment of Public Law 93-380.

Among special studies initiated at the executive level of DHEW, the recent National Evaluation of the Cities in Schools Program makes it very plain that it contains "no impact data, no measures of results" of this attempt to use schools as a base for human services delivery. The federal government's unusual interest in support of this nonagency program and its evaluation stems from interest at the Secretary level, and the emphasis on diagnosis and process seems to have reflected that level's particular interests as well.

Local Level Evaluation

No major investigation of the kinds of questions addressed in evaluation of federally supported programs at the local level has ever been undertaken. In the following remarks, we use several sources of information to characterize the activity.

UCLA's Center for the Study of Evaluation recently surveyed over 200 large school districts with research and evaluation units. The Center asked about research and evaluation generally, rather than about federal programs in particular. We have no reason to expect federal and nonfederal programs to be treated differently within units, apart from the demands engendered by reporting requirements. And so we are willing to assume that the allocation of time and priority attached to activity applies to federal government.

According to 70% of unit directors, assessing student achievement of objectives is accorded high priority and demands a substantial amount of their time. Not all these efforts are tied to a particular program, however. Of the research reports received in UCLA's survey, about 60% refer to specific programs. Checking that implementation conforms to program specification was reported to be a unit activity by about 42% of respondents and ranked high for time commitment by only 21%. Modifying programs by using evaluation results was reported to be a unit activity by less than half the respondents, with high time commitment acknowledged by only 15%. The least frequent activity of units appears to be comparing costs and benefits of alternative programs. Only 20% of the respondents acknowledge the activity and less than 2% ranked it high for time commitment.

The stress then is on tracking student progress toward goals, and to some extent, assisting in modifying the program. Answering questions about costs and benefits of alternative approaches is accorded low priority.

Our site visits to school districts were not inconsistent with the UCLA survey findings. Very few clear instances of systematic work on costs and benefits of alternative programs or of program variations emerged. Considerable attention was dedicated to testing student achievement in the interest of observing progress and making comparisons across school or district. A good deal of the work on program implementation involved interviewing teachers, principals, parents, and program staffers if such work was done at all. Differences did emerge between districts with strong research and evaluation units and those without, primarily in the range of questions addressed, production of reports, and sophistication.

We are aware of only two formal attempts to obtain survey information about how local resources are expended within a school district to answer different kinds of evaluation questions. One effort, undertaken by William Webster of the Dallas School District and Daniel Stufflebeam of Western Michigan University, focused on the 35 directors of research in the urban school districts who responded in a questionnaire survey of 60 of the largest districts. According to their responses, about 20% of resources are dedicated to answering questions about the character of program delivery and about the same proportion are dedicated to product evaluation. The remaining resources are dedicated to a wide range of other activities including management (typically around 5%), testing (typically 15%), data processing (10%). The UCLA Study of large districts with evaluation units suggests that an average of 10% of the unit's time is dedicated to meeting federal reporting or evaluation requirements. This is not unseemly in view of the fact that on average the same respondents reported that 18% of the unit's operating budget comes from federal sources.

2.5 THE DECISION TO EVALUATE

Over the past 10 years, one of the major lessons learned about evaluation is that at the national level, it is not easy, not simple, and not cheap. The demand to evaluate, however virtuous, can be agonizingly difficult to carry out. The following remarks focus primarily on the issues at the national level.

The Questions as Fundamental

Deciding which questions ought to be answered in an evaluation is fundamental. The import of this decision has been stressed in guidelines written by former Congressional staff members such as Harrison Fox and in published papers by staff of Congressional support agencies, such as the U.S. General Accounting Office and the Congressional Budget Office. It has been stressed in public papers by federal executives such as Alice Rivlin at CBO, John Evans at USOE, and Michael Timpane at NIE. It is recognized by the large school districts with sophisticated research units that we visited, and by the states with strong, if recent, tradition of obtaining sensible evidence bearing on the value of programs.

The reasons for this attention is that the questions drive all subsequent decisions, including deciding how the evaluation will be done, who will do it, and how results will be used. Answering questions about who is served may require formal information systems created by the education agency, or periodic surveys by an independent contractor, or both when there is some interest of gauging quality of the data. Questions about what kinds of services are offered may require intensive case studies or surveys, depending on how the information is to be used. Questions about what the effects or programs are may involve each of these activities simply because it makes sense to assure that somebody is indeed served and services have an identifiable character before trying to estimate effects. Determining effects of new programs on children and others generally demands more resources and planning time if estimates must be relatively unambiguous. An evaluation design must be developed and assignment of individuals to a program must accord with design. If the program is emplaced without attention to evaluation design, it may not be possible to estimate effects at all.

The questions that are asked also determine receptivity of audiences for results. The numbers of individuals served, the nature and costs of services are of interest to many managers and policy makers judging from Congressional hearings, decisions about budgets, and the like. The tradition in the United States of trying to understand systematically effects of social programs on the recipient of services is not very long. And so the audiences for these results are more difficult to identify, the debate over results is likely to be more vigorous if the conclusions are not pleasant. The decisions one can make on the basis of such information will often be debatable.

Mechanisms for Deciding Whether to Evaluate

Mechanisms for deciding whether to evaluate vary considerably across federal, state, and local levels. Generally, formal review mechanisms prevail at the federal level, and the decision process is less often formal at state and local levels.

At the federal level, the major device for making decisions about ongoing programs has been committee review of suggestions about what to evaluate and administrative procedures to support review. The approach has been taken by the Department of Health, Education, and Welfare, the General Accounting Office, and other agencies, though operating characteristics of each review group differ across agency. An analogous approach, involving committee development of a portfolio of evaluation tasks and review by the legislature, has been taken by California, according to Alex Law of the State's Department of Education. A few well developed research and evaluation units at the local level have committees to assist, review, or oversee evaluation planning, e.g., the Dallas School Board's evaluation planning committee.

Until 1980, within the USOE, the Evaluation Planning Group made decisions, within limits imposed by law and resources, about whether a program should be evaluated. No minutes of the meetings of this Group are available and no one outside government has normally been present. But the operation of the Group is traceable partly through its product, an Evaluation plan for three fiscal years which has been prepared annually, and interviews.

The first step in the process has involved annual request for suggestions, made by the Assistant Secretary for Education and made to the Commissioner, Deputy Commissioners and Executive Deputies, the Directors of the Office of Evaluation and Dissemination, and others. The Education sector activities, vested in the Evaluation Planning Group, fit into this large framework for evaluation generated by the Under Secretary of Health, Education and Welfare. Guidelines developed in 1978 and issued by the Under Secretary cover evaluation, research, and statistical activities and are detailed.

Within the Office of Evaluation and Dissemination, the response to the request involved specification of the program and project for which an evaluation is thought necessary, and the focus and purpose of the proposed evaluation.

The criteria set out to guide the submission process has included

- Expiration dates for legislation bearing on the programs and expected period in which hearing could capitalize on the information (12-18 months before new legislation).
- Programs with high priority but which had not been evaluated earlier on account of limited funds. Priority has been determined by the needs of program managers, the interests of Congress, OMB, and the general public.

- Programs in which evaluations are obsolete or otherwise invalid.

The development of a list of candidate evaluations has been an iterative process within OED as prioities, available resources, and other factors are discussed. The Evaluation Planning Group, chaired by OE's Executive Deputy Commissioner for Resources and Operations, has consisted of evaluation and policy officials. The results of the effort is a Proposed Evaluation Plan for the next three years, describing the character of the work and its costs. Final approval has been made at the Secretary level.

A major difficulty identified by agency staff familiar with the process is that the volume of work at the Assistant Secretary level has been high and the time frame too short to adequately assay the consequences of a decision. Moreover, the process is alleged to have been bureaucratically cumbersome. Part of both problems may be reduced with the creation of the new Department. We understand that a new mechanism is being developed.

Alternate approaches have been tried. For example, in the Congressionally mandated Compensatory Education Study with funds earmarked for evaluation, this committee structure was immaterial. But its equivalent had to be set up within the Study group to determine which aspects of the program or project may deserve attention. The "equivalent" amounts to a loosely defined group of individuals which include project staff and Congressional staff members with sufficient interest and ability to influence the nature of questions being addressed.

There is no regular procedure for considering whether a particular evaluation is worthwhile and the sense in which it may be worthwhile in the Congress or its support agencies. For any particular program, the procedure is rarely formalized in law. Remarkable exceptions include the recent NIE Compensatory Education Study. Among other requirements the legislation mandating that Study asked that the evaluation plans be submitted for Congressional review.

Evaluability Assessment

Evaluability assessment is a formal procedure developed over the past 10 years at the Urban Institute to facilitate the process of deciding whether to evaluate and the sense in which evaluation is possible. It asks that one first systematically "define a program in terms that agree with the manager's or policy makers' intentions," so as to permit sensible judgments what information one ought to collect and at what level of detail.

A major part of the exercise is a specification of who will use the information and how it will be used, the level of information needed to take action, and the expected impact of testing the assumptions underlying the program. The process involves considerable interaction with managers or policy makers to establish what the program is supposed to do and a model of how it is supposed to do it. Subsequent analysis is designed to understand whether the program is sufficiently unambiguous to make evaluation useful.

The process makes explicit what others do informally. NIE's Compensatory Education Survey, for instance, involved an intensive effort to get at similar kinds of information from managers and Congressional staffers before the actual evaluation was initiated. In particular, there was a formal effort to understand and document major features of program operations, notably federal, state, and local relationships, which were poorly understood at the time. It is clear that evaluability assessment is useful, however, and ought to be regarded as a legitimate procedural option in understanding whether and how to evaluate.

Until 1979 or so, most evaluations undertaken at the national level were preceded by informal, rather than formal, evaluability assessment. Current interest in making the process routine is reflected in recent activity of the Office of the Assistant Secretary for Evaluation and Program Management. A unit within the division has issued Requests for Proposals for about ten independent assessments and conducted several in-house. The in-house efforts have included an effort to better understand whether any more evaluation money could be expected to do much good in Follow Through and an assessment of the Cooperative Education Program.

The tentative policy of the new Office of Evaluation and Program Management lodges responsibility for evaluability assessment with the Division of Program Assessments, one of the units in the Management Division of the Office. Program assessments will include short term studies of three types: evaluability assessment, service delivery assessments (SDAs) and program audits. SDAs are similar in intent and practice to activities undertaken earlier by the Inspector General's Office in DHEW.

The Legislative Decision to Evaluate

There is no unique mechanism for deciding when to demand evaluation in Congress. Rather, the demand to evaluate stems from the normal process of decision-making. So, for instance, the Senate Committee on Appropriations has, during hearings, asked that OE staff obtain evaluative information for the next year's hearings. More formal requests that are incorporated into law stem at times from intensive deliberations, as in the case of the 1974 mandate for the NIE Compensatory Education Study (Elementary and Secondary Education Amendments). The requests are, at times, made law without much debate or specification, as in boilerplate requirements that the agency evaluate to determine whether the program is meeting purposes of the statute. The statutory requests may be influenced by GAO investigations, or by CBO policy analyses. To the extent that evaluation involves only case study, then perhaps this diversity is warranted. The problem of course is that level commitment may be much greater, and justifiably so. The point is that origins of the demand are diverse and that there is no special mechanism for review of demands at their source.

The absence of a special mechanism has several implications. First, it means that anyone with an interest in meeting Congressional demands can refer to no central guidance on meeting the demand well. This implies the process will be cumbersome and will demand a fair amount of interaction between Congressional staff and agency staff. It means that judgments about what can be undertaken after the fact. To the extent this process is not undertaken quickly, the opportunity for confusion increases.

To clarify the legislative decision, one may examine Reports, Hearings, and so on. These are informative for major programs and evaluations. But they are very terse and immaterial for many others. Moreover, there is no special mechanism to clarify decisions. The direct implications are that conversations about what the demand means between agency staff and Congressional staff are episodic and productive at best. At worst, they are entirely absent. The infrequency does not foster trust or at least informed skepticism necessary for a working relation. And it can lead to unnecessary suspicion. For example, we understand from Congressional staff member's public remarks that there was notable suspicion about a contractor's investigations of alternative allocation formulae in the early days of the Sustaining Effects Study. That suspicion was produced at least partly by unfamiliarity of staffers and contractors with each other. More dismal is the case of no communication between camps resulting in post facto criticism which may or may not be warranted. The absence of special clarifying mechanisms does not, we believe, help the bureaucrat respond in a timely fashion, simply because a more routine system of review also constitutes a reminder system.

Sources of Confusion. Sources of confusion are numerous. They include, we are told, the problem of educing whether the program is a new educational exercise whose effects must be determined or a civil rights mandate. The Bureau of Education for the Handicapped, for instance, regards Public Law 94-142 as a civil rights mandate: access to a free public education appropriate to their own unique needs. One can argue then that establishing that service as delivered then is a sufficient evaluation. Some educators, on the other hand, see the same law generating an education problem in that no one is well prepared to serve such children. This, in turn, may imply that estimating the effects of the program on children.

The illustration points up a more general source of conflict. There are a variety of views among federal agency and Congressional staff about whether the Congress is interested in estimating the effects of programs on the program's primary target groups, usually children. A few of the Congressional staff were emphatic in their view that Congress is disinterested. And at least one survey, by Florio and others, seems to bear that out.

The Florio et al study of 26 Congressional staffers in 1978-79 asked for ratings of the kind of information in which they were most interested. Information about effects of the program on individuals, institutions, and agencies clearly had highest priority. Costs, demographics, and opinions were clustered well below in that order.

A few agency staff were equally emphatic about Congressional dis-interest in the topic, maintaining that the primary focus should be "understanding where the money goes and who gets served." Complicating the problem is the question of how much evaluation in management's interest should be tacked onto an evaluation designed to satisfy Congressional requirement for information. The NIE Compensatory Education Study put management a poor second to Congressional interests.

The fact that statutory demands evaluations can be spawned in Congressional support agencies such as CBO and GAO, by Committees, by individual members of Congress, and others is as we've said a potential source of confusion. The diversity also serves as a rich source of ideas - and it is difficult to see how confusion can be reduced while maintaining diversity.

And of course there is lots of vagueness about decisions which might stem from the questions one addresses in our evaluation. In principle, one could specify what kinds of decisions would be made based on the information one accumulates. In practice, that specification is difficult, if not impossible because (a) insufficient time is allocated lay out decision options, (b) the nature of decisions cannot be specified well before the information is collected, (c) the decision options may change independent of the evaluation, or (d) no one is willing or able to specify decision options.

All this means that the time and effort required to clarify a simple mandate to evaluate can be demanding. Consider, for instance, the problem of developing an evaluation plan for evaluating administration of Public Law 94-142, on providing access to free and appropriate education for the handicapped. Some 18 months were required for the task and were permitted by virtue of the fact that the law became effective two years after enactment. According to Garry McDaniels and Mary Kennedy of BEH, the questions were made explicit, modified during the course of planning on the basis of interviews with members of advocacy groups, federal agency staff, and state agency staff. This effort and the law's general reference to evaluation of administration of the program resulted in questions bearing on the extent to which intended beneficiaries are served, the setting and types of service provided, administrative mechanisms in place, the consequences of implementing the law, and the extent to which the statute was met. Similarly, complicated negotiations took over six months in the NIE Compensatory Education Study of Title I.

Specifying Decisions. We have found few formal attempts to specify alternative decisions which could be made by any federal agency or by Congressional staff on the basis of a given planned evaluation. The exceptions are so-called evaluability assessments which do try to address the question of how information once obtained will be used. This precursor to a formal evaluation is demonstrably feasible at times, when the audience consists of managers in an agency. It is not clearly feasible in the legislative arena. But apart from recent studies by the GAO, it does not appear to have been tried out often in this context.

Regardless of whether decision options can be specified, there is strong disagreement about what the information implies. Consider, for example, a program found to have failed on most counts in meeting its objectives. At least one camp within the federal executive branch takes the position that because it failed, more money ought to be put into the program to make it succeed. A second camp will argue for its termination because the program has failed. Still a third camp will make the decision one way if the program is a demonstration project and the other if it was created as a service program. Complicating the matter is that similar disagreements are evident among Congressional staff. Regardless of whether decisions are specified, regardless of disagreement over implications of the data, the evaluation forms only a part of the information obtained on any program. Other information may carry considerably more weight.

This Project has not examined the decision processes carefully - our mission was to attend to a variety of other topics. The matter is pertinent here in two respects. Without prior specification of what decisions are possible if particular results emerge, determining subsequent use of results will be more difficult and may be impossible. Without prior specification it is considerably more difficult to design evaluations so as to be "relevant."

One peculiarity of our interviews with Congressional staff was reluctance of a few of them to talk to contractors who are responsible for executing evaluations. The point is pertinent here in that it may be necessary for contractors to verify evaluation goals independent of the federal agency. The reluctance was mild but vague: "A's sends contractors around to talk. . .I haven't got a lot of time. . .I am not licensed to talk to contractors by my committee chairman." The respondent who needed licensing gave the same reason for not meeting with other staffers, agency or Congressional. We do not believe this is a serious problem, but we haven't talked to a large number of staff members. If it is serious, then the prospects of clarifying objectives of evaluators are dim.

Mechanisms for Clarifying Demands and Decisions. At least one CBO staff member, believes that agency staff members do not spend enough time talking with Congressional staff, and that more time is necessary for building good evaluations by guiding better understanding of the opinions and views espoused by both groups. Similar suggestions were taken publicly in recent meetings of the American Educational Research Association by Congressional staff member John Jennings, and agency executives such as John Evans and Carl Wisler. The same spirit of exchange emerged in interviews with staff at the Assistant Secretary level.

The flaws in the system mean that evaluations are sometimes not timely and often it is not clear whether evaluations will be timely or not. Untimely reports is not typical. But the problem occurs often enough to justify concern by Jennings. The examples Jennings cites include a report on children of Title I migrant workers, which arrived too late for use in reauthorization and a request for proposal for evaluating Title IV while the program was scheduled for reauthorization in the same year. The flaws also mean confusion over what is intended by law and how the agency interprets the law and the mandate to evaluate. In public remarks at professional association meetings, Jennings cites agency confusion over how the Emergency School Assistance Act works and Congressional suspicion over the definition of income by a contractor examining Title I programs.

The elements of an improved practice appear to include:

(a) regular meetings among legislative and agency staff to make decisions about when evaluations are warranted, to determine broadly how evaluations should be carried out, and to report on progress;

(b) an information system which will make access to previous related evaluations simpler;

(c) participation by technically knowledgeable staff as well as political staff in discussion. This includes, for instance, knowledgeable staff of the Office of Evaluation and of pertinent Division of the U.S. General Accounting Office, as well as CBO and the Congressional staff.

(d) a planning system which matches production of evaluative reports to the budget cycle.

(e) planning time.

Some efforts were made in 1978-79 to remedy the problem of faculty communication, but without much success. More recent efforts include meetings at the Deputy Assistant Secretary level with Congressional staff to lay out plans for evaluation.

We believe this intention is sensible and ought to be vigorously implemented.

Footnotes

[1] In this chapter and all others, full references to the documents cited are given in the reference list, Chapter 8. The text citation includes an individual author where possible and the Section 8 lists documents by author. Where acknowledgement of individuals is not possible, the text identifies the organization that produced the report and the reference list entry identifies the organization as author.

CHAPTER 3. HOW ARE EVALUATIONS CONDUCTED?

David S. Cordray, Robert F. Boruch
and Georgine M. Pion

Always be suspicious of data collection
that goes according to plan.

In Patton, 1980.

3. HOW ARE EVALUATIONS CONDUCTED?

This chapter describes how evaluations are carried out at local, state and national levels of government. It is organized into five sections. Section 3.1 describes the basic elements of what we believe constitutes good evaluation research practice. To a certain extent, the law and regulations play a role in guiding evaluation practices at each level of government. Section 3.2 describes the explicitness with which the law prescribes evaluation methods. Section 3.3 examines similar issues pertaining to federal regulations.

Depending on the amount of discretion, resources and capabilities of an agency, federal, state, and local evaluations may exceed the requirements specified by the law and/or regulations. Consequently, describing the way programs are evaluated at each level requires an examination of the factors that contribute to evaluation practices beyond the requirements. These factors are described in Section 3.4. The last section, 3.5, takes a broader look at the type, scope and execution of numerous national level evaluations. This section provides brief illustrations of the procedures used in federal evaluations.

How well evaluations are performed depends on this material and on the capabilities of those who are responsible for their completion. The topic is discussed in Chapter 5.

3.1 ELEMENTS OF AN EVALUATION

To avoid some chronic misunderstanding here, and make plain what we believe are sensible steps in an evaluation, we describe the elements briefly. The elements are desirable, in principle, judging from guidelines issued by professional organizations, by Congressional support agencies such as the General Accounting Office, and by federal evaluation agencies. The elements include:

- Deciding to evaluate and specifying the questions to be addressed by evaluation.
- Designing the evaluation.
- Deciding who will carry out the evaluation.
- Conducting the evaluation and pertinent side studies.
- Analyzing results, making recommendations, and reporting.
- Evaluating an evaluation.
- Dissemination and use of the results.

They are not always a matter of practice.

Decision to Evaluate and the Questions to be Addressed

The decision to evaluate may be determined, as in the case of Congressionally required studies, or an agency may have some discretion in the matter. In either case, the decision should hinge on what questions ought to be addressed in an evaluation, whether anyone is interested in using the answers, whether it is possible to answer the questions in a fair and timely way with the resources at hand. If the decision is made by a legislature, the problem of defining questions is less often resolved by the legislature than by the bureaucracy asked to handle the evaluation. At the federal level and in some states, the decision may be articulated by a formal committee and checked against judgments of the legislature. If the information available for decision is insufficient and someone has the wit to recognize the fact, exploratory studies should be undertaken to obtain it. The actual form an evaluation may take depends heavily on the information available in making this decision.

Design of Evaluations

The design of evaluations depends heavily on the preceding element: it makes no sense to design an evaluation unless one knows what information is wanted, by whom, for what purpose. In the ideal case, questions are refined at the design stage of evaluation and, typically, the technical solutions to problems about how to obtain the information, at what level of detail and quality, at what cost, will be laid out. Also, in the best cases, the design stage will identify solutions to probable managerial, political-institution, legal or ethical, and scientific problems engendered by the need to evaluate. Also at its best, the process includes a review of earlier work on the topic: a literature review and conversations with those who have had a hand in producing that literature.

At the federal level, and in some states, law provides a general framework, and the task is articulated by pertinent government evaluation staff and, ideally, relevant legislative support staff. The more specific details are worked out by a contractor or agency staff responsible for specific design and actual execution of the evaluation.

Deciding who will carry out the Evaluation

Numerous classes of individuals may be designated as "evaluators." Exactly who is designated depends on the nature of the program, the level of government within which the evaluation activity is undertaken and the requirements imposed by law and regulations. These issues will be discussed in subsequent sections of this chapter and in the next chapter. Since selecting an individual outside the agency represents the most complex type of decision regarding who will conduct the evaluation, we focus here on those aspects associated with the contractor model of evaluation.

At the federal level, in education, all but a very small fraction of evaluations of ongoing programs are executed through competitively bidded contracts. A similar competitive bidding process is used in some states and at their best such states provide guidelines to local education agencies which use contractors as well. A request for proposal is issued. At its best, the request asks that the contractor submit alternative evaluation design plans and their justification if the design elaborated in an RFP is not regarded as sensible. The request should elicit fundamental information such as who will be involved, at what level of activity, at what cost, in what time frame. Milestones may be specified by either the contractor or the agency issuing the request. In the ideal case, review of proposals is based on explicit standards, laid out in the request, and is conducted by people who are well informed about evaluation from both inside the government agency and outside it.

Conduct of the Evaluation

This part of the exercise demands managerial, technical, and a variety of other skills to put plans into effect. At its best, this stage recognizes that not all problems can be anticipated and provides for side studies in budgets and resource allocation. At the national level, the demands of advisory agency such as CEIS, and of authorizing agents, such as FEDAC, or their equivalent must be met. In the ideal case, tentative clearance is given automatically and the groups actually provide useful information about the project. This stage, regardless of government level, usually involves coordinating information collection from diverse groups, school districts, for instance, public relations, management and quality control of information collection, consolidating information, liaison, budgeting and other tasks. In the simplest case, staff will be readily available and their capabilities transparent. In the more typical case, staff will have to be recruited by the project and there is a clear need to develop strategies for accommodating expected but unpredictable incompetence in at least a few individuals. This holds for both staff, advisory board members, and government staff with whom one deals. For evaluations in new arenas, pilot tests of the entire evaluation process are warranted in education as they are in health services, for instance.

Analysis, Reporting and Recommendations

Collecting information represents one phase of the evaluation process. Ensuring that the information is of high quality, reliable and valid are additional technical requirements of the process. Synthesizing the information so that it addresses and answers the evaluation questions represents the primary function of the analysis phase. Assuring the integrity of the conclusions derived from the analysis often requires careful examination and review of competing explanations. Providing additional evidence and/or a rationale for the integrity of a conclusion is often warranted in order to ensure effective communication and ward off unnecessary or incompetent criticism. Until recently, contractors at the federal level have not always been asked to provide recommendations, rather policy recommendations were derived by agency personnel. It is reasonable to expect independent sets of recommendations to be useful, providing an opportunity for input from multiple diverse sources.

Reporting in the best of cases has directed specific messages to specific audiences using a medium and style suited to each audience. In the ideal case, the clearance or review of reports is brief, the process does not prevent the report from being timely and does improve the quality of the report.

Evaluation of Evaluations

Although proper conduct of evaluation research places a premium on routine examination of the results for the presence of alternative explanations, for appropriate application of analysis procedures and for the validity of conclusions, these practices are not always conducted by the evaluator. At least one independent review and, possibly, reanalysis is warranted in many cases. Such reviews can increase the quality of the evaluation effort by (1) pointing to additional explanations which should be considered, (2) identifying questions that went unanswered in the original analysis and (3) clarifying the meaning of ambiguous aspects of the report.

Dissemination and Use of Results

Whether results of evaluation can be used and are used depends partly on what the evaluation questions were to begin with, who the audiences are, and on incentives and ability to exploit results. The organization responsible for carrying out the evaluation may also apply results, if it is linked with staff of the program under evaluation. If the evaluator is independent, as contractors are, for example, application will usually be the responsibility of others. Tracking the use of results is difficult regardless of who is responsible. That is, determining whether views are changed as a result of evaluation, whether specific decisions are made, and so on, are not easy. Evaluation forms only a part of the information available to inform any decision and separating its influence of the evaluation from the influence of pressure groups, individual intuition, and the like may be impossible. A system for routinely documenting utilization simplifies matters. This element of evaluation and the problems in execution are not much different, in principle, at local, state and federal levels.

Obstacles

The obstacles to performing any of these tasks at the local, state, or federal levels can be broadly classified into four problem areas. Managerial problems include assuring that staff are available and capable, that organization is sufficient, that cooperation from the often large number of groups whose cooperation is needed, is available, that time and other resources are sufficient. Political-institutional issues include accommodating or circumventing pressure groups, satisfying legitimate interests in short term and long term information. Scientific problems include assuring that the evaluation design is technically adequate, that the conduct of the evaluation accords with the design, that one can educe the implications of inevitable deviations from design, that one can sensibly analyze partially reliable data, and so on. Legal and ethical problems may include assuring privacy of the individual and confidentiality of response assuring due process in evaluations which must control assignment of the individual to program variations, and other matters. These difficulties are discussed in most good texts on evaluation. Research on alternative solutions to the problems is often part of major evaluation studies, but independent work is supported by federal agencies such as NIE and NSF.

Evaluation Contexts: State Administered versus Direct Grant Programs

Our discussion of the salient elements of an evaluation is idealized if one considers the evaluation process at different levels of government and across programs. There are at least three basic types of programs that can be differentiated according to the allocation process and level of government responsible for their selection and execution. These are: Direct grants awarded by the federal government to local and/or state education agencies, and two types of state administered grants--basic grants to LEAs (e.g., Title I) and special projects awarded by the state on a competitive-bid basis (e.g., Title IVc).

Evaluation is required for all programs but the relevance of various aspects of the evaluation process differ. Evaluation within the context of direct grants to LEAs (e.g., Bilingual Education) entail all of the elements described above. Evaluation of state administered programs (e.g., Title I) entails more direction from the federal government regarding how the evaluation is to be conducted, what is to be measured, how results are to be reported, and when they are to be reported. In this sense, the evaluator has less discretion over the evaluation process. Instead, the evaluation becomes a matter of first fulfilling the mandated requirements and then tailoring additional activities around these required activities -- especially when required information does not meet the needs of those individuals at each level of government. In considering how evaluations are conducted, it is necessary to consider these differences in program operation and funding. In accounting for variations in the types and quality of evaluation practices across agency levels and within the same level (across programs), federal regulations play a central role. Further, there are state-to-state differences in evaluation requirements due to state funded educational programs. These differences influence the nature and scope of evaluation practices at local agencies. Finally, within a given state, differences in the quality of LEA practices depend upon local capabilities, interest and resources.

3.2 PROCEDURES FOR EVALUATION SPECIFIED BY LAW

Some sections of the law demand that evaluative questions be addressed, but those statutes do not mention the word evaluation explicitly. Other sections mention the term evaluation but offer little guidance as to what is required. A minority of statutes are explicit as to what evaluative procedures are required. The occasional statutory references to methods, however, are interesting for their appearance at all, and suggest that it is possible to specify method in more detail if Congress wishes to do so.

In a few instances, the statutes address particular aspects of the mandated research. The broadest example of such an instance is the requirement in the General Provisions that the Secretary, in an annual report to Congress covering all federally funded educational programs, set forth "goals and specific objectives in qualitative and quantitative terms" (e.g., General Provisions for Educational Programs, Federal Evaluations -- §1226c(a)(1)(A)).

The most common specification for individual programs is the requirement that the evaluations employ "objective" measures or criteria (e.g., Title I Programs, Local Evaluation -- §2833; Federal Evaluation -- §2833(f); Adult Education -- §1207a(a)(1)(B); Emergency School Aid -- §3200(a)(11);Dropout Prevention Programs -- §3387(b)(3)). The mandate for one program provides a list of the specific measures to be included in the evaluation, including "qualitative assessments by teachers and professors, cumulative grade point average, SAT scores, acceptance to colleges and universities, college grade point average, and college major" (Biomedical Sciences Program -- §3054(a)(11)). Other mandates specify certain characteristics of the measures, such as emphasizing that the data should be comparable on a state-wide and nation-wide basis (e.g., Title I program, Federal Evaluation -- §2833(f)).

A few statutes place certain restrictions on the research design. Two mandates for evaluation specifically require longitudinal studies (Title I Programs, Federal Evaluations -- §§2833(f), Biomedical Sciences Program -- §3054(a)(11)). Other mandates offer suggestions concerning the nature of the control groups. One suggests a no-treatment control group, "composed of persons who have not participated in such programs or projects..." (Emergency School Aid -- §3200(a)(11)). Another is quite explicit in requiring an evaluation "to compare the extent to which graduates and dropouts of vocational education programs (1) find employment in occupations related to their training, and (2) are considered by their employers to be well trained and prepared for employment, except that in no case can pursuit of additional education or training by program completers or leavers be considered negatively in these evaluations. . ."(Vocational Education, State Evaluations -- §2312(b)). Another statute provided the option of conducting "not more than three experimental studies" to achieve the purposes of the evaluation mandate (Vocational Education, Federal Evaluation by the National Institute of Education -- §2563(b)(1)(D)).

Other methodological issues are addressed less frequently. Two statutes comment on the need for "statistically valid sampling" or "random sampling" in selecting the participants or programs to be included in the evaluation (Vocational Education, State Evaluation -- §2312(b); Career Education Incentive Program, Federal Evaluation --§2613(c)). Another addressed issues of generalizability by requiring that the evaluations determine if the programs have "achieved goals and are capable of achieving comparable levels of effectiveness at additional locations" (Adult Education -- §1207a(a)(1)(B)).

When warranted, explication of the type of measures, design and other conditions, in the law, clearly serves to focus the evaluation community on the informational needs of the Congress. If it is the case, for instance, that for a specific issue, the law maker believes that the use of a statistically valid sample of schools, districts, or students will provide sufficient information for their purposes, it seems sensible that the use of such a procedure should be written into the statute. Not only does the evaluator receive useful guidance as to how to proceed, there is a certain amount of uncertainty reduced as to the scope of work that is required and a sense of what the audience of the report expects to receive.

This degree of explicitness is often not warranted for a variety of reasons. In some cases, the feasibility of conducting a specific type of evaluation is unknown and the details are deferred to the agency responsible for carrying out the evaluation. In other instances (e.g., ESEA, Title I) the law requires the development and implementation of models for estimating program effects but the SEA and LEA personnel are provided with some discretion as to which model they will follow.

3.3 ADMINISTRATIVE PROCEDURES: FEDERAL REGULATIONS

The preceding section describes two functions of the legislation: authorization of evaluation at various levels of jurisdiction and the authorization of specific studies. The former type is directed primarily at obtaining information from LEAs and SEAs. The latter form of evaluation is primarily directed at national level evaluations to be conducted by independent contractors or the federal agencies. The way evaluations at LEAs and SEAs are to be conducted is typically guided by federal regulations.

Federal Reporting Requirements

Programs differ markedly with respect to the number and types of evaluative mechanisms that are described within the law and by federal regulations. To illustrate the variety of factors that affect the way evaluations are conducted, we discuss four major educational programs. These are: ESEA, Title I (Education of Disadvantaged children; basic grants to LEAs); ESEA, Title VII (Bilingual Education; basic grants and demonstration grants); Special Education for the Handicapped (P.L. 94-142, Part B); and Vocational Education (State grants and discretionary programs). These programs were selected because they are diverse with respect to their administration and organization.

This type of analysis is particularly important in that it lays the foundation for assessing whether consistent procedures are employed and for isolating where corrective action may be warranted, particularly if structurally different programs with similar information requirements reveal consistent weak areas (e.g., data quality).

(1) ESEA, Title I (Education of Disadvantaged Children, Basic Grants to LEAs)

(a) LEA Evaluation Requirements. The 1974 and 1978 Educational Amendments require the Commissioner to develop and make available to SEAs and LEAs (through the SEA) explicit standards and "models" for evaluation reporting at the local level. The October 12, 1979 Federal Register describes these standards and reporting regulations: every LEA receiving funding is required to submit an evaluation plan to the SEA that addresses how it will meet technical requirements of the regulation. At least once every three years, the LEA must evaluate its programs using "reliable and valid instruments," "procedures that minimize error" and a design that "yields a valid assessment of achivement gains." This latter requirement can be fulfilled by using one of three federally developed models or a suitable alternative approved by the SEA and Commissioner. Each model is supposed to provide an estimate of the effect of receiving Title I services on student performance compared to an estimate of what performance would have been in the absence of Title I services. Achievement scores are to be reported to the SEA using a common measure, a "normal curve equivalent" (NCE).

The new regulations also require longitudinal assessment to ascertain whether Title I gains are sustained after services are withdrawn. This assessment is for local use and reporting is not required unless requested by the SEA. Initial achievement status and gain, a description of the assessment process and project information are the only federally mandated evaluation requirements that are imposed on LEAs. The project information that is to be obtained includes: average duration of Title I service, pupil-per-teacher ratios, expenditures per child, and number of participants. According to the regulations, this project information is to be collected on a sample of grade levels.

(b) The SEA Evaluation Requirements. The SEA is charged with the responsibility for ensuring that the LEA educational plan is in compliance with the law and recently, this role has been expanded to include more extensive evaluation functions. SEAs are responsible for monitoring how the projects are carried out, providing technical assistance regarding LEA evaluation and aggregation of LEA data. The monitoring function is carried out through field visits by state Title I representative(s). The state receives one and one-half percent (set-aside) of its total allocation, or $150,000, whichever is greater, to perform these functions.

The SEA compiles the data that is submitted by the LEAs and submits (1) an annual performance report, containing: The number of participants served by types of service; number of participants by grade level for public and nonpublic recipients and "other information requested by the Commissioner" and (2) a biennial evaluation report, summarizing information for all or a representative sample of LEAs.

(c) Federal Evaluation Requirements. Section 183 of the 1978 Education Amendments clearly delineates the evaluation tasks and priorities to be addressed by the Commissioner. The law makes provision for two levels of evaluative evidence: independent evaluations designed to "describe and measure the impact of programs" and the Provision of Technical Assistance to States and local agencies on conducting evaluations. A maximum of one-half of 1 percent of the amount appropriated for these programs is provided for evaluation and priority is to be given to the federal assistance to state and local agencies.

(2) ESEA, Title VII (Bilingual Education) Evaluation Requirements

(a) LEA Requirements (Basic Grants to LEAs). Unlike Title I, these programs are direct grants awarded to LEAs or Institutions of Higher Education which apply jointly with an LEA. The guidelines for the evaluation plan appear in the rules and regulations (Federal Register, Vol. 45, (67), April 4, 1980).

As part of the application process, the grantee is requested to specify performance objectives and an evaluation plan. The proposal review procedure specifies that each proposal is rated according to 110 possible points. The specification of objectives and the evaluation plan are each allocated 15 points. The Regulations specify that the evaluation plan is reviewed for evidence that:

(1) The overall evaluation plan is consistent with the instructional training objectives;
(2) Adequate attention is paid to (a) the assessment of all objectives, (b) data collection instruments, (c) analysis procedures, (d) time schedules, (e) staff responsibilities;
(3) The design specifies a comparison procedure to estimate what performance would have been in the absence of the project;
(4) Methods to be used to identify nonparticipants for comparison or another comparison standard (e.g., an historical or statistical comparison) have been described;
(5) Sampling procedures have been identified to ensure that the sample is representative of the project population;
(6) Data collection and analysis procedures will address the evaluation questions and "are appropriate for use with the project data"; and
(7) "The data obtained will contribute to improvement in the operation of the project".

(b) SEA Requirements. The law allows SEAs to apply for "technical assistance" contracts if, during the preceding year, an LEA within the state had received funds. These contracts may not exceed five percent of the total LEA awards. Activities associated with these contracts may take the following forms: (1) Monitoring LEA Bilingual programs, (2) Evaluating the impact of programs, (3) Facilitating the exchange of information, and (4) Dissemination of materials acquired by the SEA to the LEAs. While the regulations specify that the application will be reviewed according to a point system and that the evaluation plan is to be specified, little guidance is provided as to what aspects should be considered. The review criteria for the evaluation plan is composed of a statement which simply mentions attributes such as "quality of the evaluation plan," "appropriateness of the methods" and "to the extent possible methods should be objective and produce data that are quantifiable." Except for Basic grants and demonstration projects, the same "boiler plate" statement appears in the description of the evaluation plan for all of the other program categories funded under this title. Such statements regarding evaluation components are of little use in guiding the development of an evaluation plan.

(c) Federal Evaluation Requirements. The law specifies the contents and schedule of the report to Congress. Beginning in 1980, and every two years after that, a report is to be submitted. The contents must include: a national assessment of the educational needs, the extent to which these needs are being met, a five-year plan (its costs and the needs for educational staff), and a report on and an evaluation of the activities carried out under the title.

(3) Public Law 94-142 (Education for the Handicapped) Evaluation Requirements

The Education for all Handicapped Children Act of 1975 and the pertinent regulations are explicit about responsibilities. The states are the primary target of federal oversight and they in turn are responsible for overseeing the local education agencies. The program is focused on the provision of a "free, appropriate public education for all handicapped children.[1/] The Bureau of Education for the Handicapped (BEH) in USOE was assigned the responsibility for administration and evaluation of P.L. 94-142.

(a) LEA "Evaluation" Requirements. At the local level, the term evaluation refers primarily to diagnostic assessment of children. The regulations require that preplacement evaluation be conducted using multiple, appropriate assessment modes. If the child is found to have a handicapping condition, an Individualized Educational Plan (IEP) is devised. The content of the Individual Education Plan is required by the regulations to include: (1) an assessment of present levels of educational performance; (2) a statement of annual goals and short term instructional objectives; (3) a statement of specific special education and related services and an assessment of the extent to which the child is able to participate in regular education programs; (4) projected dates for initiation and termination of services, (5) appropriate objective criteria, evaluation procedures and a schedule for reevaluation.

(b) SEA Evaluation requirements. The state has responsibility to ensure that the IEP has been prepared and that it meets the educational standards of the state. This is essentially a monitoring function and is carried out through on-site visits. Elaborate checklists have been developed by state agencies and BEH for assessing compliance with regulations. Additional monitoring requirements include fiscal audits and an assessment of the extent to which the Individual Educational Plan is actually carried out, in practice. This latter function is essentially a check to ensure that the program for individual children is actually implemented.

The law specifies that in any fiscal year, the state may use five percent of the total state allotment, under part B, or $200,000, whichever is greater for conducting required administrative activities. Evaluation in the sense of monitoring is included under this category of activities.

The State Education Agency is required to report (1) the number of handicapped children receiving services on October 1 and February 1 of the school year; (2) the number of handicapped children within each disability category; (3) the number of handicapped with each of three age groups. For all figures, unduplicated counts are required. This report is to be transmitted to the Commissioner.

(c) Evaluation requirements at the Federal Level. The Commissioner has responsibility for evaluation under Section 618 of the Act. Specifically, the legislation authorizes (1) annual studies; (2) assessment of the adequacy of information provided by state agencies; and (3) development of effective methods and procedures for evaluation.

(4) Evaluation Requirements for Vocational Education

Funding for Federal Vocational Education programs is of two basic types: Formula grants to states and Discretionary grants. The evaluation process is different for each type. Here we only consider the evaluation requirements for the formula grants administered by the states.

State administered Vocational Education programs require evaluation at the state and federal levels. At the state level, formal evaluation is routinely conducted by two groups; the State Department of Vocational Education and the State Advisory Council on Vocational Education (SACVE). At the federal level, there is a parallel organizational scheme. The Bureau of Occupational and Adult Education (BOAE) within USOE and the National Advisory Council on Vocational Education (NACVE) serve as the federal level counterparts to the state agencies. The local administration of these programs is carried out by the district. The evaluation is typically informal, being composed of needs assessment and guidance regarding program operation provided by the Local Advisory Council on Vocational Education (LACVE).

(a) Evaluation Requirements at the State level. The law and regulations are explicit as to the content and procedures to be employed in the state evaluation. The evaluation is structured around a five-year program plan. The legislation explicitly states that the purpose of the evaluation is to revise and improve the programs conducted under this plan, this plan is jointly devised by representatives of the State Department of Education and the State Advisory Council (SACVE).

State Department of Education requirements. During the five-year period of the state plan, the State Department of Education is to evaluate the effectiveness of each program in terms of (a) planning and operational processes, (b) student achievement, (c) student employment success and (d) issues related to special populations. Further, the state is required to evaluate the extent to which individuals who complete or leave the program obtain employment in occupations related to their training and whether their employers consider them well-trained and prepared for employment. Sampling is permitted for this assessment. Finally, the State Department of Education is required to submit an annual accountability report which includes a description of how funds were used, a summary of the evaluations that were conducted and a description of how the evaluation information has been used to improve the state's program.

State Advisory Council requirements. Annually, the State Advisory Council is to prepare and submit to the Commissioner and National Advisory Council on Vocational Education, an evaluation report. Its contents are to include a synthesis of its evaluation of State Department administration and operation and the evaluations performed by the State Department of Education.

(b) Evaluation Requirements at the Federal level. An organizational structure, parallel to the state level, is established within the law for the federal level agencies. There are some notable differnences in the explicitness of the evaluation requirements prescribed for the National Advisory Council, however.

Evaluation requirements for the Bureau of Occupational and Adult Education. At least ten states are to be reviewed during a given fiscal year. The purpose of the review is to analyze the strength and weaknesses of state programs. At the same time, DHEW is to conduct fiscal audits within those states. The Commissioner is to transmit to Congress a report on the National status of the Vocational Education programs. The report is to include information developed from the National Vocational Education Data System (VEDS), a summary of information obtained from federal reviews and audits and a synthesis of the evaluations performed by State Departments and State Advisory Councils.

Evaluation requirements for the National Advisory Council on Vocational Education. NACVE received a broadly stated evaluation function in the legislation. Its primary function is to provide policy-oriented annual reports and assessment of USOE/BOAE administration and operations.

Diversity in the Type of Evaluation Regulations

Examining the amount and type of information that is required across the four programs it is apparent that there are substantial differences. The direct grant type of program (e.g., Bilingual and the Discretionary grants for Vocational Education) have the least amount of oversight and reporting requirements. Title I and Vocational Education (Basic grants) are both state administered, formula allocation grants and have an additional level of evaluation imposed by the state agency. Vocational Education can be distinguished from Title I in that two agencies at the state and two agencies at the federal level are responsible for conducting routine evaluations. From this comparative assessment, we see that not only do the law and regulations indicate how evaluation is to be carried out, it can also influence how much is conducted and by whom.

3.4 HOW EVALUATIONS ARE CONDUCTED AT STATE AND LOCAL LEVELS

Although the regulations and legislation are directive as to evaluation requirements under each funding source, the way programs are evaluated at state and local levels varies dramatically, making overall statistical characterization difficult. As a consequence, types of operation are described and illustrated.

Experience derived from field interviews at six State Education Agencies and one Local Education Agency within each of these states suggests that LEA practices are influenced by educational and evaluation practices of the SEA. Attention, then, is directed first at the SEA level of evaluation.

STATE LEVEL

Until recently, the SEA was primarily an administrator of federally funded educational programs that operated in the LEAs. Mandated activities included: fiscal auditing, compliance auditing and the aggregation of data reported from LEAs. The SEA, in turn, reported to the Commissioner. With increased Congressional interest in local level evaluation, the SEAs have been given additional evaluation authority. Recent Title I legislation requires the SEA to provide technical assistance to the LEAs for the purpose of improving local evaluation efforts. Program planning and evaluation, identification of exemplary programs, and the dissemination of exemplary practices are some additional activities that SEAs have come to perform.

As might be expected, State Agencies vary with respect to their level of involvement in the evaluation of federal programs. Part of this state-to-state variation is due to the educational organization within a state and some of the variation is due to the State's own investment in educational programs. We found some states with a rather substantial monetary investment in programs that are similar to federal programs, notably State Compensatory Education and Bilingual Education. For these cases, State legislative interest in evaluation and accountability are a driving force behind the development and maintenance of a strong state evaluation component. As a consequence, state level evaluation capabilities have been strengthened, federal programs being a primary beneficiary of increased expertise. The impact of State interest in evaluation on the LEA's practices can be remarkable, resulting in many LEAs substantially going beyond federal evaluation requirements.

Types of State Education Agencies and Evaluation Practices

For simplicity, three types of SEAs are identified as a way of categorizing State level evaluation of federal programs. We offer these characterizations only as a first attempt to describe their role in the evaluation process. Little is known about the activities of SEAs with respect to the evaluation of federal programs, except of course what is required of them by federal regulations. These types are labeled: Exemplary SEAs, Compliance-Oriented SEAs and those falling between the two, referred to as Emergent SEAs. This rough classification scheme was derived as a result of our field visits, conversations with federal staff and our review of reports obtained through telephone solicitations. No statistical characterization of the prevalence of each type can be offered.

Since these categories provide a basis for distinguishing among SEAs, brief summaries of their distinctive features are provided. These descriptions are not intended to be complete portrayals of each type, they are merely thumbnail sketches which summarize, roughly, the discrepancies found among SEAs with respect to the way federal programs are assessed.

(a) Exemplary SEAs. These are characterized by an active interest in the evaluation process at numerous levels of state organization. These levels include: the State legislature, the program sector and the evaluation component. Common elements among states of this type which seem to be responsible for this interest include: administrative or legislative fiscal support for the state level evaluation component, few administrative levels between the evaluation component and the Chief State School Officer, substantial monetary investment in State funded educational programs and strong public support for educational accountability. From an evaluation perspective, perhaps the most salient characteristic of this type of SEA is the presence of a well articulated evaluation plan which encompasses state and federal programs within the program planning, implementation, evaluation and dissemination processes. The impact of this overall plan is seen throughout the SEA, across program areas and it permeates the LEA level; influencing evaluation practices and program administration.

(b) Compliance oriented SEAs. The absence of one or all of the characteristics identified for Exemplary SEAs is likely to influence staffing levels, monetary support and/or technical capacity (e.g., computer facilities) that make it simply impossible to go beyond the minimal requirements established by federal regulations. As a consequence, efforts are, by necessity, directed at ensuring that the minimum standards are adhered to and there is little opportunity to do anything else. For the purpose of federal reporting requirements, a compliance-only mode of operation should not be viewed negatively so long as the quality of what is reported is high or at least the level of quality is known. In these cases, the role of the Technical Assistance Centers is likely to be extremely important.

(c) Emergent SEAs. Recent developments in several arenas have contributed to improved evaluation practices at State Departments of Education. These factors include: federal development of specific guidelines pertaining to evaluation requirements, Technical Assistance in the implementation of these guidelines, increased availability of trained evaluation personnel, and direct federal assistance designed to strengthen SEA capabilities. SEAs classified as "emergent" are those which are no longer simply complying with regulations. Instead, agency-initiated practices are developing with respect to how evaluations are to be carried out for their own needs.

Emergent SEAs can be distinguished from the exemplary category on a number of dimensions. A primary distinction is the extent to which the exemplary practices are exhibited across a variety of programs. On this dimension, an answer to the question how or how well are evaluations conducted would require an answer, prefaced by the statement, "it depends on which program you are referring to." An additional distinction that might be drawn between these SEAs and the exemplary variety pertains to the institutionalization of the enterprise -- evaluation in the emergent states is only now beginning to gain credibility.

Illustrations of State Level Evaluations

Taking into consideration the discussion of elements comprising the evaluation process described in Section 3.1 and federal requirements described in Section 3.3, the most salient issues pertaining to how evaluations are conducted at the state level can be summarized as follows:

1. Program planning and approval;
2. State on-site monitoring: compliance and program reviews;
3. Specification of reporting requirements;
4. State level analysis, aggregation and data quality control.

The extent to which each of these elements is addressed within state level evaluations is the focus of this discussion. Illustrations which highlight the distinctions among the three types of SEA will be made wherever possible.

(1) Program planning and program approval. The legislation and regulations assign responsibility to the state for approval of LEA applications for program funding and the evaluation plan. This function is particularly relevant for state administered programs. Two basic strategies were observed in the SEAs we visited: (1) "on-paper" review to ensure that the program as planned is consistent with federal regulations and (2) compliance review plus an "educational-quality" review. Examination of local and state reports suggests that this latter form of planning is more typical of exemplary SEAs. However, there are notable legislative and regulatory exceptions across programs that attempt to induce better planning. The legislation, for example, pertaining to Vocational Education requires the state agency to explicitly indicate how the previous year's evaluations have been used to improved program operation. A similar requirement within Title I applications was observed in a few sites. California employs the notion of a Master plan in order to achieve continual improvement of educational practices. In support of this effort the SEA disseminates written material on program improvement and conducts site visits. In California, this aspect of the educational planning process is part of the development of a Consolidated Application plan which allows LEAs to devise programs relevant to individual LEA needs. Title I and four state funded programs are involved in the California's Consolidated Application process. Our interviews with State Department personnel suggest that this coordinated effort results in better targeting of funds for specific student needs.

The more common planning strategy appears to be a review process where programs are first examined "on-paper" to ensure they meet federal standards. At this point, the state either approves the program or recommends modifications. In well established programs, such as Title I, the submission, approval and funding process can become routinized, resulting in little substantive change over time. Consequently, the same goals, level of attainment and the like could be specified from year to year.

Regulations and legislation provide for a certain amount of flexibility as to how programs can be operated. This discretion translates into numerous choice points regarding the operation and instructional characteristics of local programs. Our experience suggests that in many cases these choices are made in the absence of systematic tests as to the impact of one alternative or an other. We found little attention being directed at the state level towards systematic tests of alternative means of delivering program services. Such alterations could be easily designed into the program planning process and systematically examined. The feasibility of introducing systematic tests of program alternatives for ongoing programs into the planning process can be illustrated with a couple of examples.

Example A: Testing program components. Title I has regulations specifying that funds can only be used for the purpose of supplementing, not supplanting, regular education efforts. In the plan submitted by LEAs, the manner in which this regulation is to be satisfied is specified. This typically involves the selection of a particular educational method (e.g., pull-out). Given the state's oversight responsibility for approving such organizational/operational schemes, there appears to be a considerable amount of room for SEAs to encourage LEAs to systematically examine the impact of differing project level operating schemes. In our site visits, we did not encounter the use of this type of planning/evaluation strategy.

Example B: Program variations. Length of exposure and mode of instruction, independent of program settings, represent additional aspects of program operation which could be examined rigorously by SEAs provided adequate attention is paid to methodological issues during the planning phase. Through coordination with the State Agency, different "dosages" of exposure-time could be allocated to a representative sample of capable LEAs in order to assess the impact of exposure on achievement.

By virtue of the state's oversight responsibility for approval of the educational and evaluation plan to be carried out by the local agency, urging the SEA to select those LEAs who have adequate resources to participate in these types of systematic tests of alternatives seems sensible.

(2) On-site monitoring: compliance and program quality review. In the evaluation literature, it is well known that what is planned is not always the same as what is ultimately implemented in the operational field setting. The administrative analogy of this principle is compliance monitoring. The regulations specify guidelines pertaining to compliance and SEAs are responsible for ensuring that the LEA adheres to them, in practice.

States differ with respect to the amount of on-site monitoring that is conducted. In addition to the fact that states have varying degrees of monetary investments in programs similar to the federal programs, the federal allocation for administration, of which monitoring is a component, is poorly structured. A set-aside of the total state allocation with a fixed ceiling limit has been designated within the law. The amount of the set-aside is sufficient in some cases but strongly favors those states with few LEAs. States with a large number of districts, many with 300-500 districts receiving some federal funding, find it difficult to monitor the troublesome sites let alone all the sites. Of course, representative sampling of sites is sufficient for monitoring purposes to obtain a statistical characterization of compliance but if the issue is to obtain an outside check on the extent to which programs are implemented as planned, "spot-checking" would be insufficient.

An exemplary instance of a program-quality-review procedure is offered by California. In addition to examining program compliance, quality of the school program is assessed through site visits using the Program Quality Review Instrument (PQRI). As part of this review process, the California State Department of Education recently conducted an assessment of this review process by soliciting opinions of those individuals who participated. Similar follow-up procedures were instituted by the Vocational Education evaluation personnel, this assessment was designed to ascertain the extent to which recommendations prescribed by the site review team were adopted by the local agency.

Well defined compliance monitoring systems were observed for numerous programs across SEAs that were visited. In many instances, state procedures are more extensive in scope than federal regulations prescribe. For example, compliance monitoring for 94-142 is coupled with more stringent compliance regulations of State Programs in New Jersey and in Massachusetts, though each employs somewhat different procedures tailored to state legislative requirements.

The importance of compliance monitoring, especially coupled with program quality review cannot be understated. Current administrative procedures are such that, more often than not, the evaluation of local level programs is insufficiently funded to allow outside consultants to perform these audits. However, given the fixed set-aside allocation for program administration under certain title (e.g., Title I, Bilingual), similar staffing problems are likely. This appears to be an area where more attention is needed.

An admirable use of Bilingual Title VII funds at the State level for improving the monitoring and ultimately the reporting practices of LEA grantees was observed in Massachusetts. A series of studies was commissioned by the State Bilingual Office to examine state Title VII programs and their evaluations. One result of these investigations was the development of contract specifications, reporting guidelines and standards for Bilingual evaluations. Further, the plan developed by the outside contractor was scheduled to be pilot-tested prior to full-scale implementation. From a planning perspective, these practices are admirable and should be promoted. Similar, thoughtful planning and monitoring was observed in other program areas in the Massachusetts SEA.

(3) State specification of reporting requirements. Judging from recent legislation, Congress has a considerable interest in obtaining program information and outcome evidence on a nationwide basis. This is not a new concern. The distinctiveness of the recent legislation is that it represents a more direct request and, in many cases, substantial monetary backing is provided (e.g., Title I). From the Federal perspective, a major concern is that data collection and aggregation procedures are comparable across states -- the primary federal goal is national-level aggregation. The SEA is ultimately responsible for ensuring that federal mandates are fulfilled. These data collection and aggregation procedures are common tasks ascribed across program areas for state administered programs. The ultimate utility of this evidence at the Federal level is dependent upon state-to-state consistency regarding what is collected, in what format and by whom.

States and programs within states vary considerably as to how reporting requirements are fulfilled. Most states rely on a two-phase aggregation process. Typically, the LEA collects project level information and performing the first level of aggregation. The State then receives LEA reports performing the second aggregation.

The possibility for error is compounded at each level of aggregation and the ability to account for distinctive features of a particular project becomes more difficult. The necessity for coordination within and across states is evident from the recent flurry of activity associated with the development of information reporting systems in Compensatory Education (TIERS), Vocational Education (VEDS), and Education for the Handicapped through BEH.

In an effort to understand the issues related to obtaining high quality data from LEAs, current data collection procedures for Title I reporting were examined. Specifically, the contents of Title I Annual Evaluation Reports for 1978-1979 submitted to 10 different states were examined. Table 1 provides a summary of the information that is required of LEAs under the auspices of these "Evaluation" reports. In Table 1, an "X" designates that this information is required, a "0" indicates it is not even mentioned and a "?" signifies that it is impossible from visual inspection to ascertain whether the item is required or at the discretion of the LEA. In specifying the possible content that might appear in these reports, Title I requirements and elements that were characterized earlier as part of the evaluation process were used as the basis for comparing state evaluation reporting schemes. As such, the comparison across states is devised to include information on characteristics of the program, parent advisory council, staff, testing and other desirable evaluation components. Examination of the pattern of entries in Table 1 across states for each information element, reveals a substantial amount of overlap on what is reported. For example, the testing cycle (spring to fall, spring to spring, or fall to fall) is clearly identified in all 10 State forms. Also, the reviewer can easily discern whether scores are reported for only students with both before and after scores, the grade level is always specified and the program area (Mathematics, Reading, etc.) is easily detectable. Further, Title I staff characteristics and Parent Advisory Councils characteristics are almost uniformly reported.

Areas where there is less consistency across states include test identification, explanation for the discrepancy between the number of students served and the number tested, and method of scoring the test. Each of these factors is important in the aggregation process in order to understand the quality of the data. Of substantial policy relevance is the identification of program characteristics. It is well known that programs and their modes of operation differ. We see that program setting is often omitted in these reports. Some program settings involved "pulling" Title I students out of their regular classroom, others are self contained classrooms. Student exposure to the program is another variable of interest, and as late as the 1978-1979 school year it was not uniformly reported. This variable can be used to illustrate yet another level of complexity that must be considered when data are reported from the LEA. Although the exposure level was reported by many states, closer inspection of the way it was reported reveals considerable variation. For example, it appears that some LEAs reported the amount of exposure that they had *intended* to provide each student,

Table 1

Evaluation information required in LEA 1978-1979 Annual Evaluation Report for Title I Programs (Regular) for Ten States

Type of Information Required	States									
	1	2	3	4	5	6	7	8	9	10
Programmatic Information										
Number served										
unduplicated counts	x	x	0	x	x	x	x	0	x	x
duplicated counts	x	x	0	x	x	x	x	x	x	x
Subject Area	x	x	x	x	x	x	x	x	x	x
Program setting (e.g. pullout)	x	0	0	x	0	0	0	0	x	0
Student Exposure	x	x	0	x	x	0	x	0	x	0
Staff/student ratio	x	0	0	x	x	0	x	0	x	0
Expenditures	0	x	0	x	x	x	x	x	0	0
Parent Advisory										
Composition of committee	x	x	0	x	x	?	x	x	x	x
Activities of committee	x	x	0	x	x	?	x	x	x	x
Title I Staff	x	x	0	x	x	x	x	x	x	x
Staff inservice	x	x	0	x	x	0	x	x	x	x
Test Information										
Testing cycle	x	x	x	x	x	x	x	x	x	x
Complete test identification	x	x	x	?	x	x	?	?	0	0
Method of scoring	0	x	0	0	0	0	0	0	0	0
Matched on Pre/Post scores	x	x	x	x	x	x	x	?	x	x
Explanation for missing scores	x	0	0	x	?	?	0	?	0	0
Test Results										
Reported "on level"	?	x	x	0	?	x	?	?	?	?
Reported by grade level	x	x	x	x	x	x	x	x	x	x
Reported by school, project	0	0	0	0	0	0	0	0	0	0
Reported in NCE's	x	x	0	x	0	0	x	x	x	x
Additional Elements										
Narrative with interpretations	0	0	0	x	?	x	?	0	0	0
Recommendations	0	0	0	x	?	x	x	0	0	0
Assessment of other objectives	0	0	0	?	0	x	x	0	?	0
Dissemination/Feedback	0	x	0	0	x	?	x	x	0	0

others seem to have provided estimates of the actual amount of time the children received services. If the issue is to identify those programs that are exemplary, consistency of reporting seems essential. This level of detail is important and a lack of specificity is certainly a source of confusion for those who must comply with reporting regulations. The new Title I reporting system (TIERS) is designed to provide explicit indicators of program characteristics which will improve the degree of consistency within and across states; its quality remains to be seen, however.

The last few rows of Table 1 show that there is considerable diversity in these reports, across states, with regard to the inclusion of narrative summaries, recommendations, assessment of objectives other than achievement and dissemination/feedback. These characteristics were isolated earlier as important aspects of the evaluation process, however they are given little attention in state reporting requirements. As a consequence, the "evaluation reports" are in many cases merely summaries of head counts, aggregate test scores and little more.

(4) State level aggregation and quality control. The aggregation process can take basically two forms: (a) translation of what is received into common units and summation or (b) in addition to these tasks, further analyses/evaluation may be undertaken. The Michigan, California, and New Jersey State Departments of Education represent instances of the latter category. Michigan used data from its 1975-1976 Title I evaluation report to examine program effectiveness by examining building-level programs within types of school district. California's assessment of Title I (and State funded programs) included the use of multiple achievement tests, program quality reviews obtained through site visits, and quality assessments of the data reported from LEAs. Similarly, New Jersey regularly assesses and reports its evaluation of Title I and the State Compensatory Education Programs simultaneously.

More typical of the state reports that were reviewed is aggregate reporting with varying degrees of attention to data quality. Notable exceptions are those State Departments who explicitly list the amount of data that has been excluded, the reasons for exclusion and the bias that is likely to result due to their exclusion from the aggregate analysis. Further, this type of careful data management has been found to be useful in identifying LEAs which should be targeted as candidates for technical assistance.

Conducting evaluation of federal programs through the aggregation of data elements from diverse sources requires that careful attention be directed at the quality of the collection process. This is an expensive proposition. It is not sufficient to simply provide data forms that give the appearance of generating compatible data. Close monitoring of the process is mandatory. The Census Department regularly conducts validation audits. This practice is similar, in principle, to the required audit of Individualized Education Program plans (IEPS) undertaken by BEH. Consideration should be given to the idea of routinely assessing the validity of information that is reported to the SEAs. BEH, in preparation for the implementation of reporting requirements associated with Public Law 94-142, commissioned an assessment of state counts of handicapped children and produced a handbook for conducting future validation studies. Systematic sampling of LEAs seems to be an economical and sufficient means of assessing the quality of data elements included in reporting systems. However, given the diversity in types of data collected under various titles, it is likely that audit procedures tailored to the specific data system will be necessary.

HOW EVALUATIONS ARE CONDUCTED AT THE LOCAL LEVEL

Despite the specificity of regulations, there exist LEAs who perform remarkable evaluations. Unfortunately, when we view these in the context of the 16,000 LEAs who participate in federal education programs, these exemplars seem to be the exception rather than the rule. Even in the large districts that were site-visited, mixed levels of practice were observed -- some merely fulfilled evaluation requirements and others went substantially beyond.

Existing Evidence on How Evaluations are Conducted Within LEAs

Catherine Lyon and others, at UCLA's Center for the Study of Evaluations (CSE), identified LEAs with enrollments in excess of 10,000 students and an evaluation/research unit. Among other issues, Lyon and others were interested in characterizing types of evaluation activities carried out within these units. Directors of the research unit were asked to complete a questionnaire pertaining, in part, to the activities and relative amount of time the unit devoted to these activities. Roughly 230 Directors responded to the questionnaire.

The CSE data is interesting in that it provides a rough indication of the extent to which local districts, with evaluation units, perform certain types of activities. Further, since each activity was ranked according to the relative amount of time it consumer, a crude characterization of the methodological emphasis within these units can be devised. Table 2 reports pertinent data from the CSE study.

Examination of Table 2 reveals that all districts collect information on student achievement and 95% of the Directors listed this category as one of their three most time consuming efforts. Given the emphasis on achievement testing in the regulations, this is not too surprising. Information on the relationship between student and/or classroom characteristics and achievement are less likely to be collected, nor are they ranked as being time consuming; nearly 60% reported that they did not engage in collecting this information. The relationship between socio-economic status and achievement is collected in fewer than half the districts. With respect to time allocations, these activities were ranked as being one of the three most time consuming activities by only 26% of the Directors.

Table 2 also shows that data collection by means of testing is almost universal (98.8%), followed by survey questionnaires (93.8%). The prevalence of interviews and classroom observations is considerably lower (65.2% and 60.4%, respectively). Turning to the evaluation activities performed by evaluation/research units, it can be seen that there is high variability among

Table 2

Activities, Functions and Methods Employed by Evaluation/Research Units at LEAs: CSE Results

Evaluation Activities	Percent of Districts	Ranked 1,2 or 3
Collection of information: (12 original items)		
1. Student achievement	100%	94.7
2. Relationship between school/classroom characteristics and achievement	41.5	25.6
3. Relationship between student achievement and socioeconomic status	49.3	17.6
4. Relationship between students' race/ethnic background and achievement progress	40.1	16.3
Methods of Data Collection: (5 original items)		
1. Testing	98.8	94.2
2. Survey questionnaires	93.8	89.0
3. Interviews	65.2	43.2
4. Classroom observation	60.4	37.9
Evaluation activities: (11 original items)		
1. Assess the results or worth of instructional programs	88.1	69.6
2. Assess student achievement of objectives	91.1	66
3. Identify/appraise educational goals or objectives	77.1	34.4
4. Compare the districts' achievement test scores with scores outside the district	81.0	27.7
5. Determine pupil and/or public satisfaction with school or programs	65.1	22.4
6. Check that implementation conforms to program specification	63	20.7
7. Modify programs using evaluation results	63.8	14.5
8. Approve evaluation sections of program proposals	71.8	11.9
9. Assist in the selection of instructional programs	45.9	7.1
10. Compare costs/benefits of alternate instructional programs	22.4	1.3

districts as to the activities performed and the relative amount of time devoted to each activity. Assessing the results or worth of a program and examining the achievement of objectives are carried out by at least 88% of the district, each being ranked as one of the three most time consuming activities by about two-thirds of the Directors. Comparing district results with other districts, assessing satisfaction with the school or program, approving evaluation plans, and modifying programs using evaluation results is a relatively frequent activity in that at least two-thirds of the Directors claim to perform these efforts, yet they are not often ranked within the three most time consuming activities. Further, even though checking to ensure that the program is implemented as planned is a crucial phase of the evaluation process, 37% of the Directors report not engaging in this activity. For those who do, it is ranked as one of the three most time consuming activities by 21% of the Directors. Analysis of alternative instructional programs is conducted within 23% of the districts and is ranked as a time consuming activity by less than 2% of the directors.

Webster and Stufflebeam examined how evaluations are carried out within evaluation units in urban school districts, by obtaining estimates of budget expenditures for a variety of evaluation activities. The Webster-Stufflebeam analysis is based on thirty-five urban school districts categorized according to size of the evaluation, research and testing budget: one million dollars or more (large), $300,000 to $999,999 (medium), and $50,000 to $299,999 (small). The percentage of budget expenditures for each of their classes of activities appears in Table 3. In addition to the average percentage within each district size category and weighted average across district size, Table 3 presents the range of values composing each average. The ranges are interesting in that the presence of a 0 as a lower value indicates that none of the budget is expended for that activity.

Roughly thirty to forty percent of the budget is expended on testing and assessing whether the program is successful in meeting its objectives, or is more successful relative to an alternative method of instruction. Providing evidence regarding implementation of the program consumes an additional ten percent of the budget, on average. The same is true for data processing. Further, providing assistance to other district personnel (research consultant), proposal development, providing ad hoc information and instrument development, combined, constitute roughly 20% of the budget. Needs assessment and diagnosis of constraints upon meeting needs (context evaluation) and evidence regarding availability/use of resources (input evaluation) represent less than 7% of the budget.

Another important consideration in describing activities in LEAs is the degree of diversity among LEAs. For example, testing, in the aggregate represents less than 20% of the budget expenditures. Examining the ranges, we find as little as two percent and as much as 70% expended on testing. Similarly, product evaluation (success, impact of the program) commands as little as 5% to as much as 50% of the budget. Of particular interest are the 0% values. Across districts, there are some districts (within each category) that do not allocate any money for input evaluation, the same is true for the research consultant and proposal development categories. There are some medium and small districts who do not allocate much (if any) of their budget to assessing the implementation of the program.

Table 3

Percentage of Budget Expenditures for Evaluation Activities:
Webster and Stufflebeam (1978)

Evaluation Activities	District Size			Weighted Average
	Large (9)	Medium (13)	Small (13)	
1. Product Evaluation (Assessment of the relative success of the program in meeting its objectives relative to an alternative method of instruction, cost/benefits)	19.0% (6-50)	19.6% (10-40)	17.9% (5-50)	18.8% (5-50)
2. Testing (Operation of the system to ensure quality instrumentation and reporting).	11.8 (2-22)	23.4 (4-70)	19.0 (5-45)	18.8 (2-70)
3. Data Processing (Actual operation of basic information system).	13.0 (5-25)	7.6 (3-16)	10.2 (5-15)	10.0 (5-25)
4. Process Evaluation (Providing information on factors affecting implementation and for aiding the interpretation of program evaluation data).	14.0 (5-25)	10.9 (1-20)	5.5 (0-10)	9.7 (0-20)
5. Management of evaluation related resources.	5.8 (3-10)	8 (4-19)	10.7 (5-18)	8.4 (3-18)
6. Ad hoc information (Provision of requested information on an ad hoc basis).	3.8 (0-15)	4.4 (2-10)	10.8 (3-20)	6.6 (0-20)
7. Context Evaluation (Assessment of needs, description of outcomes, actual and desired diagnosis of problems that prevent needs from being met).	6.8 (2-18)	6 (2-15)	5.3 (0-13)	5.9 (0-18)
8. Planning services (Technical assistance in planning and managing projects or programs).	4.6 (0-15)	4.4 (2-10)	4.9 (1-11)	4.6 (0-15)
9. Research Consultant (Assistance in design and analysis of projects conducted by other district personnel).	3.6 (0-12)	4.7 (0-10)	4.3 (0-8)	4.3 (0-12)
10. Proposal Development (Development of proposals or the evaluation sections of proposals for outside fundings).	3.9 (0-11)	4.3 (0-10)	4.0 (0-10)	4.1 (0-11)
11. Other research (Basic and applied).	6.9	2.4	3.8	4.1
12. Instrument development and validation	3.7 (1-12)	4.1 (2-15)	4.0 (0-10)	4.0 (0-15)
13. Input Evaluation (Providing information on resource availability and utilization for accomplishing goals).	3.1 (0-6)	.2 (0-3)	.6 (0-5)	1.1 (0-6)

The CSE study and the Webster and Stufflebeam study, while focused on different aspects of how evaluations are conducted at the large local districts, reveal a common theme -- the primary mode of evaluation at the local level is directed at assessing whether a program has met its objectives and a considerable amount of time and money is devoted to testing as a means of describing project success. Fewer resources are devoted to process assessment, though it receives more financial support than needs assessment type activities and assessment of the adequacy of resources for meeting goals/objectives. Further, there is substantial variability in the way districts allocate resources and the amount of time expended on various types of evaluation activities.

Both of these studies approach the question of how evaluations are carried out from a research unit perspective. Little is known about the distribution of activities for specific programs. Recalling the diversity of evaluation activities specified in the legislation and regulation across programs, it would be expected that not all of the evaluation activities would be equally relevant. Based on the data presented above, it would be inaccurate to characterize any particular evaluation as being composed of 20% impact assessment, 20% testing, 10% monitoring and so on. To obtain a better understanding of how particular evaluations are conducted, it is necessary to examine what procedures are employed for specific programs. Unfortunately, little data bearing on this issue is available. Some survey data are available, however.

The National Center for Education Statistics recently examined the methods employed by LEAs for the evaluation of Title I. In particular, they focused on how frequently each of the three OE Title I Models were used. Using a national probability sample, it was estimated that 87 percent of all districts had Title I programs in 1978-1979. Of those districts with Title I programs, 63 percent used an evaluation model. Of those using an evaluation model, 90% voluntarily followed one of the three OE evaluation models. The remaining 10% used a local developed model. Each OE model entails before-after testing of students, they differ as to what form of comparison is employed. The simplest OE model entails using the publisher's norms as the basis for comparison. This model was used by 86% of those districts who used any model. The remaining 4% was split between the control group model (Model B) and the special regression model (Model C) developed by OE, both of these require testing students who did not participate in Title I. Technically these latter models are considered to be methodologically more sophisticated, providing less equivocal estimates of gain that can be attributed to the program. The NCES survey shows that models employing comparison groups are infrequently used. The predominant mode of achievement assessment is through comparison with norms, that is, what a sample of non-Title I students would have achieved given their initial level of performance.

Across the three studies, there appears to be a substantial emphasis on testing as a means of evaluating programs. The NCES survey shows extensive use of the before-after norm-based assessment. The use of other methods is infrequent and NCES estimated that districts would require additional assistance in developing more complete evaluation practices (e.g., continuous program improvement, selection of non-test based evaluation devices).

Reporting requirements for many programs emphasize achievement testing. As a consequence, it seems reasonable to expect that fulfilling minimum requirements, for example, before and after testing of only those students who receive Title I services, would be the most frequently observed. However, while alternate methods are used infrequently, the fact that 400 districts were estimated to be using some form of comparative assessment suggests that they are feasible and their use should be promoted.

Another glimpse at how evaluations are conducted can be obtained by examining the procedures used by projects that have been approved by the Joint Dissemination and Review Panel (JDRP). Each year a brief summary of the approved projects and the evidence supporting its effectiveness is published in a volume entitled Educational Programs That Work. The most recent edition contains 34 newly approved programs; methods employed for these most recent programs were examined. Specifically, of the 34 programs listed, 62% employed comparison group designs, 56% employed standardized tests or a combination of standardized and locally developed tests and roughly 30% involved some form of replication where the program was evaluated more than once, revealing consistently positive results. For approved exemplary projects, the most common design entailed comparison of project performance where one group did not participate in the project or program. It is also of interest to note that 38% of the evaluation didn't employ comparison group designs. The most typical being a before-after assessment of the project participants only. In many cases the evidence supporting the achievement gain was supplemented with additional evidence that the program was sufficiently implemented; that the gains could be replicated from year to year; and that the use of multiple testing devices revealed consistent evidence.

Illustrations of Evaluation Practices at LEAs.

Up to this point, the discussion has attempted to provide statistical characterizations of how evaluations are conducted on LEAs. We have made reference to the fact that multiple procedures are employed in an individual evaluation effort. To illustrate the process of conducting evaluations at the local level, case studies of five evaluations encountered during the site visits are provided. Material for the last case study appeared in a recent publication.

> Case I. In site D, the evaluation unit and the instructional programs are administratively and fiscally independent. To avoid miscommunication during the course of an evaluation, a team composed of a member of the evaluation unit and a program specialist from the district's Instructional Programs Division is assigned to each school. The team assists the school personnel in drafting and measuring their program objectives which ultimately become a component of the application for funding. This approach was devised to ensure that advice regarding curriculum and issues pertaining to evaluation of the program were resolved by appropriate experts.
>
> For program development the team negotiates with the school personnel in forming appropriate objectives; a second team reviews the plan as a form of external validation. In the planning process, previous evaluative

evidence is considered. This evidence is provided by the evaluation unit and entails an individaulized report which summarizes the extent to which previously stated objectives have been attained. Test results are summarized in graphic form so as to facilitate communication. Through periodic monitoring of the classroom acitivies, process information is collected and systematically reported to the school personnel.

Case II. The evaluation of Title I in site C is best described as compliance-oriented. The evaluation unit produces a report that is in accordance with SEA reporting requirements.

To augment the minimal testing information provided by the evaluation unit, the Title I program staff engage in a variety of evaluative activities that they initiate and conduct, independent of the evaluation unit. For example, one component of the program--a kindergarten language development program-- was assessed in terms of language skill achievement scores. The testing and data analysis was done by program speech therapists and the program coordinators. Technically, this report was not very sophisticated; inferences as to the effects of the program were overstated. Other examples of locally initiated evaluations conducted by program staff include an assessment of parent and project staff needs and an evaluation of one of the inservice workshops. Both of these assessments were carried out through the use of locally developed questionnaires or modifications of existing instruments. In the absence of cooperation from the evaluation unit, the program staff in this site resorted to conducting their own evaluations.

Case III. The evaluation of Title I in site G is similar to that observed in Case I, above. The major exception is that, while the evaluation unit is administratively independent of the program, they are not fiscally independent of the program. The Director of Federal Programs is responsible for the operation of Title I in this district, evaluation money is also distributed by this individual.

Within the district evaluation office, a full-time evaluation specialist is assigned to the Title I evaluation. Close association with the project staff results in a cooperative working relationship. The evaluation specialist monitors the progress of the program at each site through formalized classroom observations, test results, and monitoring of the completion of specified instructional objectives.

To ensure that the evaluation remains objective, the State Department of Education, "strongly suggested" that an outside contractor be hired to review the evaluation documents produced by the district's unit.

Case IV. In site D, the evaluation of the Title VII Bilingual Education program was conducted by an outside contractor. The evaluation was monitored by the research/evaluation unit and the program staff. Prior to acceptance of the final report, the program staff and the evaluation personnel requested that additional analyses be performed. In this way, through technical assistance from the evaluation unit, the program staff were able to make certain their informational needs were satisfied and that the evaluation information was valid.

The working relationship between the evaluation unit and the program staff is also depicted by their collaboration in the development of entry and exit criteria for participants involved in the Bilingual program.

The evaluation process in this district regularly involves a combined effort of outside contractors and evaluation unit staff, A similar arrangement was observed in sites G and B.

Case V. In site E, the evaluation unit is responsible for assessing compliance with the regulations for P.L. 94-142 (Education for the Handicapped). Information is gathered through parent, citizen, community agency surveys, analysis of a random sample of Individual Education Plans (IEPs), and teacher/administrator surveys.

At the request of the program director and through additional funding (about $5000), an examination of the operation and cost effectiveness of service delivery was undertaken. This evaluation entailed an analysis of the amount of time devoted to program tasks, the number of students served across numerous schools, and ratings of the quality of service that was provided. Document analysis and survey procedures were employed for the quality of service assessment. An alternative placement procedure was also examined.

Case VI. Paul Rost (1980), from the Albuquerque Public Schools, reported on the use of a four-component planning, evaluation and dissemination system that attempts to facilitate the use of evaluative information for local planning of Title I. A manual, referred to as an *Item Documentation*, was developed to provide teachers with an indication of what each test item was designed to measure. By reporting test results on an item-by-item basis, the teacher, through the use of the *Item Documentation*, can readily determine which skills should be focused on for each child.

A *School Report* synthesizes evaluation data in a manner that makes it conducive to planning. The *Resource Allocation Plan*, based on Title I needs assessment data, includes a summary of the school's relative need for supplemental assistance. Finally, the *Title I Impact Review* is conducted. Here, the instructional team reviews the evaluation data with a member of the evaluation staff. The instructional team provides additional interpretation of the evaluation -- giving reasons for specific outcomes based on events occurring during the school year.

Evaluation of Title I in Small and Large LEAs

With the exception of the NCES survey of school district's use of the USOE Title I models, most of the available evidence on evaluation within LEAs, focuses on districts with large enrollments or those with evaluation/research units. Nationally, there are nearly 16,000 operating districts; less than five percent with enrollments in excess of 10,000 pupils. Since our site visits included large LEAs, a broader basis for judging how evaluations are conducted was established through a telephone survey directed at LEAs in four categories: enrollments of 2500-5000 pupils; 5000-10,000; 10,000-25,000; and those with 25,000 or more pupils. LEA personnel responsible for the administration of Title I were interviewed and follow-up discussions with specific evaluation personnel were undertaken whenever possible and/or warranted. The interview questions focused on how evaluations were conducted, who is responsible for their completion and how are the results used.

In describing how evaluations are conducted, it is useful to establish how much of the Title I award is devoted to evaluation and the administrative placement of evaluation within the LEA. Information pertinent to these issues is summarized in Table 4. To assay the amount of federal money used in the evaluation of Title I, LEA personnel provided information on the amount of their most recent Title I award and the amount devoted to the evaluation effort. The estimates derived from these responses show that, across LEAs, about 1.5% of the Title I budget is devoted to evaluation. The smallest percentage for evaluation -- .8% was reported by districts with 2500-5000 pupils. In terms of actual dollar values, this is a trivial expenditure for evaluation. The most frequent comments offered by officials in these LEAs were that (1) costs incurred for evaluation were assumed by the district, (2) the evaluation was based on a pre-existing testing program which add no extra cost, or (3) only testing material were purchased out of Title I funds. Of the remaining LEA clusters, the average set-aside ranged between 1.6% and 1.8%. For large grants, these set-asides are sufficient to hire multiple full-time personnel. Using the percentage of the grant award is a convenient way to summarize the extent to which federal money is allocated to evaluation but does not take into consideration the size of the grant, nor the diversity across LEAs. Figure 1 provides a graphic display of the diversity among LEAs with respect to the actual amount of money devoted to evaluation across Title I awards. The most striking aspect of Figure 1 is the diversity of money allocated within each award category. Second, if we consider the actual dollar figure allocated across districts, only about 38% of the districts reported allocating $10,000 or more of Title I funds for evaluation. Roughly 51% of the districts report spending in excess of $5000 of Title I funds for evaluation; 16% reported spend no federal money for evaluation.

In many cases, small LEAs reported that "filling out the forms only took a few days" so they found little reason to allocate any funds to the "evaluation". We found little consensus among LEAs as to how much of the budget could or should be devoted to evaluation.

Table 4 also summarizes information on the administrative placement of evaluation within LEAs. Consistent with our experience from the site visits, in the majority (69.4%) of large districts, the evaluation personnel do not report to the program officials directly, that is, they are administratively independent. As the size of the district decreases, the extent to which evaluation is independent of the program drops to roughly 50%; in the smallest districts in our sample, less than one in four are administratively independent. If we consider independence in terms of who has control over evaluation expenditures, the degree of independence accorded the evaluation is further reduced across all categories of LEAs -- 41.7% of the evaluations in large LEAs are fiscally independent of the program; only 14.8% are fiscally independent in smallest LEAs in our sample. Considering administrative and fiscal independence together, the values diminish even further. If we judge the import

Table 4

LEA Title I Evaluation Allocations and Organizational Arrangements: Telephone Survey Responses

LEA Category (Enrollment)	(A) Number Contacted	(B) Response Rate	(C) Percent for Evaluation	(D) Administratively Independent	(E) Fiscally Independent	(F) Both D & E	(G) Beyond Regulations
25,000 and above	40	100.0% (40)	1.6% (38)	69.4% (36)	41.7% (36)	36.1% (36)	86.4% (37)
10,000 to 25,000	38	94.7% (36)	1.7% (32)	45.5% (33)	21.2% (33)	21.2% (33)	80.0% (35)
5,000 to 10,000	38	86.8% (33)	1.8% (33)	50.0% (28)	21.4% (28)	17.9% (28)	75.8% (33)
2,500 to 5,000	34	82.4% (28)	.8% (27)	22.2% (27)	14.8% (27)	11.1% (27)	39.3% (28)
	150	91.3% (137)	1.5% (129)	48.4% (124)	25.8% (124)	22.6% (124)	72.2% (133)

Note: The number of usable responses for LEAs in each cluster appear in parentheses.

Figure 1

LEA Dollar Allocation for Title I Evaluation for Different Size Title I Grant Awards: FY 1979-80

Dollar Allocation for Evaluation from the Title I Award	Amount of ESEA, Title I Grant Award								
	$100,000 and below	$100,000 to 250,000	$250,000 to 500,000	$500,000 to 750,000	$750,000 to $1 Million	$1 to 2 Million	$2 to 5 Million	$5 to 10 Million	$10 Million and More
$0	8	5	3	3	1	1			
0-500	7	5	1						
500-1000		4	1						
1000-2000		1	3	1					
2000-3000	1	1	3	2	1	1			
3000-5000			4	2	4				
5-10,000	1	4	6	3	1	1	1		
10-20,000		1	3	3		2	1	1	
20-30,000			2	3		7	1		
30-50,000			1		1	3	3	2	
50-100,000				1		1	3	2	2
100,000 +							2	5	

Source: Telephone survey.

Note: The dollar amounts are unverified reports from LEA personnel responsible for Title I, Education for Disadvantaged.

of these figures from the context of our discussions with LEA personnel, the lack of fiscal and administrative independence does not usually result in constraints on how and how well evaluations are conducted, there are noteworthy exceptions to this generalization, however. These constraints are addressed in Chapter 5.

With respect to how Title I evaluations are conducted, two major issues are pertinent: the extent to which consistent procedures are employed across LEAs and the extent to which LEAs supplement the required evaluations in order to satisfy their own needs. With regards to the first question, 85.3% of the reported using one of the three USOE Title I models, the most predominant (80.1%) being the norm-referenced Model A. Consistent with the NCES survey, only 5 percent of the LEAs used the Model B, C, or an approved alternative design -- these are most often conducted within large school districts.

The second issue relevant to how evaluations are done is whether and in what ways LEAs supplement the required Title I evaluation. The last column of Table 4 shows that with the exception of small districts, at least three-quarters of LEAs go beyond federal reporting requirements to some degree; less than 40% of the small districts (2500-5000 pupils) report supplementing beyond federal requirements. A follow-up question on the types of supplemental evaluation activities that are undertaken revealed that a variety of strategies are employed. The more sophisticated districts perform side-studies to assess the validity of achievement data, assess implementation of the program, and to determine the extent to which the program environment is different than the comparison group's educational environment. Also, elaborate feedback mechanisms have been developed to provide evaluative information to teachers, parents and other local officials. More typical supplemental activities that were reported include: one-time surveys of teacher attitude, assessment of changes in the affective domain produced by exposure to inservice workshops or the Title I program, classroom observations, and assessment of locally developed material. Judging from material received through site visits and telephone solicitations, the quality of these additional activities ranges from poor to exceptional. However, for a substantial number of districts with enrollments of 5000 or more, there appears to be at least some effort devoted to obtaining information that is suited to the needs of the district personnel.

3.5 HOW EVALUATIONS ARE CONDUCTED AT THE FEDERAL LEVEL

The preceding section addressed the evaluation of federal programs at the state and local levels of government. These evaluations represent one aspect of how federal programs are evaluated. The second source of evaluative data is derived from national level studies initiated by the federal agency or explicitly mandated by the Congress. This section focuses on the procedures associated with the latter type of evaluative studies.

Evaluation within OED

The Office of Evaluation and Dissemination at USOE engages in numerous evaluation activities that can be classified into five major categories. These include: (1) Impact studies designed to measure the effects of programs on participants; (2) Process studies to provide information on the characteristics and operation of programs administered by USOE; (3) Technical Assistance in evaluation to State and Local Education Agencies; (4) Identification of exemplary projects and educational practices; and (5) Exploratory Evaluations aimed at specifying the goals and objectives of OE programs. In this section, we focus on the way evaluations are conducted in OED. The primary focus is on an examination of the procedures undertaken in carrying out impact and process studies.

After the decision to evaluate: The execution of an evaluation.

Within the context of why evaluations are conducted, a detailed description of OED mechanisms for deciding which programs will be evaluated was provided. Selection of the studies to be undertaken depends on the priorities established within OE, constraints imposed by mandates, reauthorization cycles, and resources. Once priorities have been established, procedures are initiated within OED to facilitate execution of these studies.

The essential phases of the OED procedures are as follows:

Phase 1. An OED staff member is assigned to the study shortly after it is established as a high priority and usually remains the project monitor throughout the duration of the study.

Phase 2. The monitor and program staff and additional evaluation experts convene a series of meetings to discuss the nature and scope of the issues to be addressed by the study. An evaluation plan is negotiated at this phase. The plan includes specification of the questions to be addressed, objectives of the study, and methods to be employed. The feasibility of the evaluation plan is considered at this phase through consultation with experts in the field and evaluation personnel within OED.

Phase 3. Having specified the scope of the study, negotiated an evaluation plan, a Request for Proposal is written. These RFPs are explicit as to procedures that are to be used in conducting the study.

Phase 4. A contractor is selected through a competitive bid process.

Phase 5. Once the contractor is selected and the evaluation is initiated, the OED project monitor maintains close contact with the contractor throughout each aspect of the study. The basic purpose for this monitoring function is (1) to make sure that the study is conducted in accordance with the original plan, (2) advise the contractor when problems arise which may jeopardize the validity of the study, and (3) ensure that study, as executed, answers the intended questions.

Phase 6. At the completion of the contract, the monitor takes responsibility for the production of a project summary and ensures that the report is reviewed and approved by the Department and ultimately transmitted to the appropriate audiences.

Types of Questions addressed and characteristics of OED Studies.

In discussing why evaluations are conducted, the type of evaluation questions that are addressed served as a functional classification scheme. In Chapter 2, four categories of questions were offered, entailing five specific questions. These questions include:

1. Who is served?
2. How are services provided?
3. What are the costs of these services?
4. What are the effects of these services?
5. What are the costs or benefits of alternative forms of service?

As reported earlier, the nature and cost of service are the most frequent questions addressed in federal level evaluations. It was also noted that multiple questions are addressed in many instances. Since the nature of the evaluation effort is dependent upon the type(s) of questions that are posed, it is useful to look at the complexity of these studies with respect to the number of questions that have been addressed.

Applying the classification scheme devised in Chapter 2 to the sixty-four (64) studies reported by OED during 1977-1979, we find that a single question is addressed by 45% of the studies; two questions are addressed in 37.5% of the cases; and three or four questions are addressed in 14% and 3% of the studies, respectively. The major inference that can be drawn from these figures is that, for the most part, OED evaluations are focused on rather specific evaluative questions.

It was also indicated earlier that contrary to what had previously been claimed, only a minority of the studies undertaken by OED were focused on obtaining answers to questions pertaining to the impact of educational programs. Specifically, over the past three years, 25 of the 64 completed studies were focused on impact questions; that is, answering questions regarding the effects of services on program participants. Of the twenty-five impact studies, only 16% (or 4 of 25) were directed at simply looking at the impact of service. Instead, for 84% of the studies, multiple questions were considered. Specifically, 36% answered the impact question and one additional question; 36% addressed impact and two other questions; and four of the five questions were addressed in 12% of these impact studies. None of the OED studies attempted to gather evidence pertinent to all five questions.

So, while the majority of OED sponsored studies are focused on answering one or two specific questions, when we focus attention on just those studies that are aimed at understanding the impact of programs on participants, it is apparent that these studies are rarely focused on answering only the impact question. Incorporating diversity as to what is examined provides information which can serve the needs of multiple constituencies not only those who are interested in the impact. Further, coupling questions regarding impact with assessments of the nature and/or cost of service provides a more complete understanding of the reasons for the observed level of impact.

Characteristics of OED Studies

Who conducted these studies? Closer inspection of the 64 evaluations reported in the OED Annual Reports reveals that 60 specific studies were conducted. Numerous groups were involved in the execution of these studies. Of the 60 studies, 5 were conducted by OE staff, 1 by NIE, and the source of 2 studies could not be readily identified. The remaining 52 studies were conducted by individuals outside OED. Specifically, 30 different groups participated in the conduct of these studies; 22 research firms were used, 2 universities, 2 federal agencies other than OE and 4 organizations classified as "other."

How long do these studies take to complete? The duration of each contract can serve as a rough approximation as to the length of time that is required to complete an evaluation. The range of the time devoted across the OED 52 studies, conducted outside the agency, is presented below.

Distribution of Contract Duration for OED Evaluations

Months	Percent of Contracts	Cumulative Percentage
73 or more	1.9%	100.0%
61-72	3.8	98.1
49-60	1.9	94.2
37-48	7.7	92.3
25-36	25.0	84.6
13-24	36.5	59.6
6-12	11.5	23.1
less than 6	11.5	11.5

From the distribution it is apparent that there is considerable diversity in the duration of contract--the shortest being only one month and the longest exceeding 6 years. More typical durations, however, are in the range of one to three years. Nearly 85% of the contracts are completed within three years, the average being 24.5 months and a median duration of 22 months.

How much do evaluations cost? Because of the confusion over what is and what isn't considered to be evaluation research, it is exceptionally difficult to obtain an idea of how much is spent conducting national level studies. Even though the 52 contracts described here do not represent a scientifically based sample of evaluations, examination of the amount of money allocated can be informative as a bench-mark as to the rough costs of conducting national evaluations.

The gross, unadjusted dollar values allocated to each contract are categorized below, yielding the following distribution.

Dollar allocation (Total=$52,003,061)	Percent of contracts	Cumulative percentage	Percentage of total allocation	Number of contracts
$2,000,000 and above	9.6%	100.0%	57.7%	5
1,000,000 to $2,000,000	11.5	90.4	16.4	6
750,000 to 1,000,000	7.7	78.8	6.9	4
500,000 to 750,000	5.7	71.1	3.4	3
250,000 to 500,000	21.1	65.4	7.9	11
100,000 to 250,000	23.1	44.2	7.0	12
50,000 to 100,000	7.7	21.1	.4	4
10,000 to 50,000	9.6	13.5	.2	5
below 10,000	3.8	3.8	<.1	
			100.0%	52

Examination of the distribution reveals at least two major classes of national studies. The first may be considered as "large scale" studies, those costing in excess of one million dollars. These are a minority, accounting for roughly 21% of the contracts. For this class, the maximum allocation was slightly over $17 million dollars for the 5 year Sustaining Effects Study conducted by Systems Development Corporation. More typical for the class of evaluations is an allocation of less than $3 million. The second class of national studies is reflected in the substantial concentration of awards on the $100,000 to $500,000 range, accounting for about 44% of the contracts.

Given that the average duration of a contract is on the order of two years and most (65.4%) are allocated at less than $500,000, the impression that all national level studies are large-scale, expensive undertakings, does not appear to be substantiated. On the other hand, if percentage of total dollar allocation ascribed to each category of contracts is considered, the large-scale studies that were conducted account for nearly 75% of the money allocated for the 52 studies -- the Sustaining Effects Study, alone accounts for one-third of the $52 million dollar expenditure. While expensive, large-scale studies still represent only a fraction of the activity undertaken by OED.

The studies being considered were initiated in the early to mid 1970's. The focus of evaluations conducted through OED has changed since that time;

large and medium-scale efforts are less often initiated than in the past. Evidence for this change in orientation appears in the form of increased emphasis on exploratory/evaluability assessments and increased expenditures for provision of Technical Assistance.

Two illustrations of OED evaluations. Statistical characterization of evaluations can be informative in that crude parameters can be established as to how long they take, what questions are addressed, the number of questions that are addressed, and so on. However, beyond these global descriptions, there are considerable differences among national studies. The complexity of the issues addressed and the type of evidence that is required inevitably requires an evaluation design that is tailored to the specific questions. This tailoring concerns multiple aspects of the evaluation process. Namely, measures must be selected that are appropriate to the objectives of the program; when interviews are conducted, target groups need to be identified and questions developed that are pertinent to each group. Design considerations concerning the allocation of individuals to the program conditions must be specified or the evaluator must identify comparison groups, after the fact and so on. The number of procedural details that ultimately contribute to the quality of the evaluation are almost impossible to enumerate meaningfully. There are some general themes, however.

To illustrate the commonality and variety of methods that are likely to be encountered, brief summaries of salient procedural aspects of two recent OED sponsored studies are provided as Exhibits A and B. Exhibit A summarizes the Vocational Equity Study conducted by American Institute for Research. Exhibit B depicts the procedural aspects of the ESSA, Magnet School Program Study carried out by Abt Associates. Both studies illustrate a number of common procedures employed in national level studies; their uniqueness is also apparent.

As for the common elements: Both studies have multiple objectives; A sample of schools, individuals, or other unit is examined; and information is obtained from multiple sources -- in both cases, information from previously existing sources and new data collection procedures are employed.

On the other hand, there are notable differences in the technical aspects of each study. The Vocational Education Equity study used a scientifically based probability sample in order to meet the intent of the legislation. Here, the legislation requested an assessment of sex in equity in all vocational education programs. The contractor, noted that such a wide-scale assessment would not be feasible given the number and diversity of program sites. The sample was designed to estimate the extent of inequity in all programs. Within schools, the individuals who were interviewed were also drawn according to sampling procedures. These are expensive procedures.

For one phase of the study, the identification of exemplary projects, the sample was not scientifically based. Rather, deliberate, purposive,

EXHIBIT A

VOCATIONAL EDUCATION EQUITY STUDY

CONTRACTOR: AMERICAN INSTITUTE OF RESEARCH, Palo Alto, California

AUTHORS: HARRISON, DAHL AND OTHERS

INITIATED BY: CONGRESSIONAL MANDATE

PURPOSE: Section 523(2) of the Education Amendments of 1976 (P.L. 94-482) directs the Commissioner of Education to conduct a "study of the extent to which sex discrimination and sex stereotyping exists in all vocation programs assisted under the Vocational Education Act of 1963." The mandate also requests an assessment of the progress that has been made to reduce or eliminate these practices.

OVERVIEW OF PROCEDURES: Eight objectives for this study were identified by staff at the U.S. Office of Education. These included: Assessing the extent of sex discrimination and stereotyping in vocational education programs and the progress which has been made in reducing sex inequities; General progress which has been made towards reducing sex inequities; Identification and analysis of practices and activities that hinder or facilitate equal opportunity at state, local, and school levels, as well as external factors; To identify and analyze programs that are successful at reducing sex bias; and to develop criteria by which federal, state, and local agencies can measure progress towards equal opportunity.

The study conducted by AIR entailed three primary evaluation components:

- Primary data collection
- Secondary data analysis and literature review
- Case studies of selected programs

SCOPE OF WORK:

- Interviews with State Agency personnel in 49 States and the District of Columbia.
- 100 schools, offering vocational training, were drawn in accordance with probability sampling techniques.
    - At each school, 4 Counselor, 8 Teachers, and 35 Students were selected according to probability sampling techniques. Background characteristics of each respondent, attitudes and perceptions of sex inequity were obtained through structure interview procedures.
- Secondary data analyses were performed by a sub-contractor using data from a variety of sources. These included: the Census Bureau, the Office of Civil Rights, Bureau of Labor Statistics, and Enrollment data obtained by the Bureau of Adult Education (BOAE).
- A nationwide search for exemplary and promising projects that attempt to reduce sex inequity was conducted by contacting experts in the field, a survey of relevant literature, and from nominations obtained through site visits. Twelve sites were visited and documented as case studies. A review of the available evidence, documenting the projects' effectiveness, served as a criterion for their being included as a case study. Twelve other sites were described as promising and given a brief writeup in the report.

FUNDING LEVEL: $957,968

DURATION: 24 months

EXHIBIT B

STUDY OF THE EMERGENCY SCHOOL AID ACT MAGNET SCHOOL PROGRAM

CONTRACTOR: ABT ASSOCIATES, INC., Cambridge, Massachusetts

AUTHORS: EUGENE C. ROYSTER, D. CATHERINE BALTZELL, FRAN CHERYLL SIMMONS, AND KAREN FERB

INITIATED BY: USOE

PURPOSE: This study had four major objectives:

(a) To describe ESAA magnet schools, and other comparison magnet schools, including characteristics of the school districts;
(b) To assess the overall effectiveness of the magnet school as a desegregation device;
(c) To assess the conditions for a successful magnet school program; and
(d) To examine the operation of the ESAA Magnet School Program.

OVERVIEW OF PROCEDURES:

The study entailed 3 primary evaluation activities

- primary data collection
- analysis of administrative records pertaining to school enrollments
- review of documents

SCOPE OF WORK:

- 18 school districts were selected, at each district, at least three magnet schools were visited by two members of the research team. Site selection was based on characteristics of the school and from these categories, districts were drawn at random. The site visits took 3-4 days each. All procedures were pilot tested in three sites. In total, data were obtained from 227 magnet schools.
- Data collection included:
  (1) unstructured open ended interviews with district administrators, principals, teachers (in magnet and nonmagnet schools), parents, representatives from community groups, and school board representatives;
  (2) quantitative data included school enrollment reports, 1970 census for demographic data, desegregation plans for court orders.
- Effectiveness of the magnet schools as a desegregation device required the construction of two indices, subscription rate and racial balance. These were derived from enrollment data, projected enrollment data, for minority and majority students. The criterion for success was established as a subscription rate of 1.0 or more. Racial balance was determined as the difference between majority and minority subscription rates. Longitudinal assessment of district desegregation, in magnet and non-magnet schools, was undertaken using a common Index of Dissimilarity, Interracial contact and a ratio of the two indices.
- Factors correlated with magnet school effectiveness were isolated.
- Project selection, criteria for review and program operation were also examined.

FUNDING LEVEL: $258,827

DURATION: 16 months

sampling was undertaken to find exemplary practices. This type of purposive sampling is common practice in national studies that attempt to identify exemplary programs. As a procedure, it is well suited given the intent of the study -- it is not adequate for estimating effects, however.

In the ESAA, Magnet School Study, districts were categorized according to general characteristics and then randomly selected from each category. The major concern here was that the selected districts were representative of the variety of contexts and conditions that Magnet schools operate. Since this study was not designed to estimate the extent to which all ESAA Magnet Schools impact on desegregation, the attention to sampling of interviewers at each site was not relevant, as it was in the Vocational Education Equity Study.

Description of the procedures employed in national studies, that go into relentless detail, are not warranted in a report of this type. However, the important aspect of this discussion is to point out the necessity of matching the procedural aspects of the evaluation with the nature of the question. Estimating the overall effects of a program or policy is technically more complicated than other types of questions -- as a result they are more costly. Thus, the importance of the clarity with which the question was asked can drastically impact the level of effort devoted to a particular study.

Evaluation in the Bureau of Education for the Handicapped

The Education for All Handicapped Children Act of 1975 (P.L. 94-142) was not put into effect until October of 1977. According to Mary Kennedy, the delay between the time the law was passed and its eventual effective date allowed for the development of an evaluation plan encompassing agencies at each level of government. The core of the evaluation plan revolves around an analysis of the information needs of various audiences and stipulations in the law. Six major questions ultimately emerged from the BEH assessment of informational needs. Since 1976, 20 Special Studies have been initiated or conducted, all of them are or have been conducted by outside contractors. Each study has addressed one or more of the six basic questions identified by BEH staff. The percentage of studies that have been conducted for each question appear below.

1. To what extent are intended beneficiaries being served?	40%
2. In what settings are they being served?	20%
3. What services are being provided?	25%
4. What administrative mechanisms are in place?	20%
5. What are the consequences of implementing the law?	60%
6. To what extent is the intent of the law being met?	65%

The percentages do not total 100% because multiple questions were addressed by many studies. Eleven of the 20 studies addressed two issues, 6 addressed only one issue and three studies span aspects of all 6 questions.

Of particular relevance, is the type of information obtained through these studies. Prior to the implementation of the law, a series of assessments was conducted to ascertain state capabilities to respond to reporting requirements. Since the State counts, depicting the number of handicapped children and their condition(s), are the basis for allocation of funds, a study was initiated to determine the validity of these counts and methods for validating future counts were formalized. Next, assessment of state definitions was undertaken in order to account for state-to-state variations in the number of handicapping conditions that are reported. In addition, an assessment of the difficulty of implementing Individualized Educational Programs was undertaken. Each of these studies provided information that was ultimately used in formulating policy for reporting requirements.

Subsequent efforts have included analysis of data obtained from states, a five-year study of a sample of school districts in which progress in implementing the law is being observed, and an 18 month study of the first year of implementation at nine local school systems. The Congress, in the legislation, mandated a national survey of the nature and quality of individualized educational programs, this three-year effort was initiated in 1977.

In 1978, a series of 5-year longitudinal case studies to ascertain the consequences of P.L. 94-142 on various participants was initiated. Each of these is funded at between fifty and sixty thousand dollars per year. Five studies were initiated to address specific issues associated with the implementation of the law and its impact on learning disabled students, parent activists, quality of educational services, differential impact for students with various handicapping conditions, and school-parent relations.

The BEH Evaluation plan. The foregoing describes the Special Studies component of the evaluation of P.L. 94-142, other sources of evidence are used in the overall evaluation of P.L. 94-142. Kennedy describes how individual sources of evaluative information are incorporated into the overall evaluation. The basic concern was to ensure "breadth" (national coverage) and depth, or rather, an understanding of the processes associated with the implementation of the law. Four sources of information comprise the overall evaluation of P.L. 94-142. These include: (1) descriptive data from state agencies; (2) data collected through federal monitoring of 17 specific provisions of the act; (3) independent surveys and (4) case studies and other small-scale focused studies. Whereas the descriptive data obtained from the state provides national coverage of the effects of implementing 94-142, the special studies are intended to complement these data in that they provide relevant information as to the process of implementating the law and its impact on the quality of the educational process.

This overall plan represents an admirable use of multiple sources of data which minimizes the reporting burden placed on SEAs and LEAs, yet provides relevant information for understanding the impact of implementing the law.

CHAPTER 4. WHAT ARE THE CAPABILITIES OF EVALUATORS?

Georgine M. Pion, David S. Cordray,
and Robert F. Boruch

"The evaluators...were spread throughout...each endowed with the special gift of their own group, and each using that gift in a special way.... But does that not make for much arguing among evaluators about who has the most special gift of all?"

In Patton, 1980.

4. WHAT ARE THE CAPABILITIES OF EVALUATORS?

Chapter 4 presents a preliminary examination of the capabilities of evaluation performers. The first section addresses some of the misconceptions associated with this topic. The next two sections are devoted to enumerating the problems associated with the identification of evaluators and ascertaining their capabilities. In Section 4.4., the institutional context and resources for evaluation are described with respect to their influence on capabilities. Evaluation tasks, capabilities, and the match between them are outlined in the fifth section for evaluators in federal, state and local education agencies. Given that the use of outside contractors has recently aroused much interest, the final section explores some of the salient issues involved in the procurement, monitoring, and capabilities of outside contractors.[1]

4.1 MISCONCEPTIONS OF EVALUATORS AND THEIR CAPABILITIES

There have persisted various misconceptions concerning evaluators and their capabilities. A common one has been that of "the evaluator" who supplies all the necessary skills for any evaluation effort. Much attention has been devoted to describing the characteristics of this "person." While this may accurately depict those situations in which one individual is contracted or assigned to conduct the research, there are many instances where a group of individuals share evaluation responsibilities. In these cases, judging the capabilities of each individual against all tasks required by the evaluation is inappropriate. Rather, capabilities of the group as a whole should be compared against the necessary tasks.

In addition, the idea of capabilities needs to be more broadly construed to include not only those of the primary evaluation staff, but also the various resources which they have at their disposal. For example, access to such facilities as Technical Assistance Centers, interested universities, and other agency departments with trained personnel must be considered when determining the presence or absence of evaluation capabilities. At the same time, such factors as money and staffing patterns cannot be overlooked. These may affect the organization's ability to attract and retain competent individuals.

Another misconception is that there is a detailed armentarium of talents necessary for each and every effort associated with evaluation. This view does not take into account the possibility that such efforts as evaluability assessments may not require exactly the same set of competencies as do cost-effectiveness strategies. Responsibility for evaluation reporting to federal/state education agencies may not require the ability to plan, design, and conduct a program evaluation. Consequently, there is no immutable set of specific competencies against which all individuals responsible in some way for educational evaluations can be judged.

4.2 PROBLEMS IN IDENTIFYING EVALUATORS

In education, several of the same problems in identifying "evaluators" surface as do in other arenas (e.g., mental health and criminal justice). For example, anyone can easily assume the title of "evaluator." There is no formal licensing procedure and no long tradition of training or certification. The problem here is akin to the identification of engineers in the 1930's and to some extent now--the school coach might be labeled "evaluator" in the same sense that a boiler mechanic is labeled "building engineer."

Identification of who is and who is not an evaluator, let alone "competent." cannot simply be achieved through the use of one explicit indicator. First, individual job titles often inadequately denote the presence of evaluation responsibilities. While in some SEAs and LEAs there is explicit mention of "evaluation" in the position title (e.g., Evaluation Specialists and Evaluation Technicians), in other settings these individuals are known as Divisional Assistants or Educational Research Specialists. Part-time doctoral students, occasionally employed to perform evaluation tasks, are typically referred to as hourly workers or temporary help. Even the inclusion of departmental or divisional affiliation is not instructive since many agencies do not have a distinct evaluation office. In these settings, individuals function in multiple roles--program administrator, principal, and evaluator--and, if queried, would more closely align themselves with the administrative or instructional title and would not regard their tasks as evaluative.

No single characteristic such as evaluation training or experience can identify all evaluators. Formal degrees or certification in evaluation are rare, given the recency of the field, and graduate programs are new and vary substantially in the type and quality of training they provide. Adoption of the title of "evaluator" by an individual also does not necessarily imply that he/she possesses the appropriate qualifications or has actually conducted an evaluation. The recent recognition that evaluation is a growing enterprise with employment opportunities may have exacerbated this problem. We are reminded of a conversation with an architect who, following the surge of interest in social impact assessments, simply appended "evaluator" to the specialties listed on his business card. In education, including oneself on a state registry of available "evaluation consultants" can be accomplished through a simple phone call.

Because a large number of people are often involved in an evaluation, the problem of whom to include becomes important. Here the question concerns whether or not the principal investigator, the project director, the field coordinator, the instrument developer, or the analysis manager all equally qualify as "evaluators." Time allocations and types of evaluation activities cannot be rigidly assigned to these roles. For example, the principal investigator(s) for major contracted evaluations may assume full-time responsibility for the evaluation or simply serve as a part-time "mentor" or fiscal manager. Even targeting only the author(s) of the evaluation report can be problematic, given that authorship is not always explicitly stated.

Highlighting the problems associated with identifying evaluators is not to denote the futility of the effort but only the complexity which is introduced into discussions of evaluators and their capabilities. One cannot simply decide to look at only those individuals explicitly labeled as "evaluators" by their job title or employment in an "Evaluation Office" as this would exclude a number of potentially relevant participants. It cannot also be assumed that only one particular individual conducts the evaluation and therefore is "the evaluator." Reliance on registry listings of "evaluators" or "consultants" may result in the inclusion of individuals who have never been involved in educational evaluations. We have tried to be sensitive to these issues in our research and examine those people who currently have major responsibility for evaluation efforts related to federal education programs. These activities include large-scale national evaluations of federally funded education programs, evaluation reporting to federal/state agencies, and discretionary evaluation efforts related to these programs at the state and local levels.

4.3 SPECIFYING THE CAPABILITIES OF EVALUATORS

In examining the capabilities of evaluators, two questions emerge: (1) what are the capabilities required for evaluation? and (2) how can these be understood? Concerning the first question, many individuals in academia (e.g., Anderson and Ball), private contracting firms (e.g., Ingle and Klauss), and urban school districts (e.g., Webster and Stufflebeam) have attempted to delineate the skills necessary for evaluation. However, even these experts have often failed to reach perfect agreement. Aside from crude differences in the numbers of skills and terminology, there also have been differential emphases attached to various competency areas. While the majority of individuals have included such competency areas as managerial and communicative styles, technical proficiency has typically received the greatest attention. In contrast, inclusion of capabilities related to programmatic skills (i.e., substantive knowledge in the program area being evaluated), policy-related expertise, and "credibility" has been less uniform. Enumerations of the specific levels of proficiency needed for various skill areas have also differed. Given the early state of development of the profession, this diversity is not surprising.

Factors Requiring Recognition in Specifications of Capabilities

What has resulted from this variation, however, has been conflicting images of the evaluator. One describes a technically sophisticated expert, facile in all evaluation and research methodologies. Another which is commonly evoked by some texts and short training sessions suggests that anyone can become an evaluator after surveying a "cookbook" of evaluation methods or attending a three-day workshop. This confusion stems from a number of factors, but most importantly, from differing assumptions of the tasks involved in evaluation. For example, the model of the "lone evaluator" requires a wider range of competencies within one individual than does the notion of "an evaluation team" where individuals are assigned a limited number of specialized tasks. Above all, the tasks required for simple evaluation reporting to federal/state agencies do not warrant the same degree of proficiency in many competency areas as does engaging in discretionary evaluation efforts which go well beyond compliance with federal/state regulations. Consequently, capabilities must be matched to the evaluation tasks which are assigned.

Pragmatic considerations cannot be overlooked. Evaluation resources, in terms of funding, time, and staffing levels, play an important role. Organizations operate within visible constraints--if positions are not allocated for evaluation, it is impossible to hire any staff, let alone persons with evaluation training and experience. If evaluation funds are miniscule, it becomes difficult to recruit individuals with sophisticated backgrounds, given the cost of professionally skilled labor. The lack of nearby universities or centers with evaluation expertise makes competent advice, however brief, difficult to obtain. Such issues suggest that these factors cannot be ignored in discussions of evaluation competency.

Gratuitous Criticism

Ignoring factors such as the ones specified above often results in less-than-constructive criticism of evaluation capabilities. In judging capabilities, it is difficult to resist judging competency based on what activities one thinks should be occurring vs. what activities are required. For example, districts and state education agencies who merely but competently engage in required evaluation reporting activities cannot be assumed to be incompetent or even incapable of engaging in additional discretionary efforts--they may be simply prohibited from doing so. Instances may exist where componentwise testing and assessment of program variations are impossibilities, given inadequate funding and administrative support.

Given these conditions, we believe that rather than an ideal evaluator, there is only an ideal evaluation. The elements of "good evaluation practice" have been set forth earlier in this report. The ability to successfully implement and execute these tasks rests partly on the individuals involved and partly on the available resources. Both of these will be addressed in this chapter.

Indicators of Capabilities

With these considerations in mind, assessing capabilities requires the simultaneous use of several crude indicators. First, relying on formal training in evaluation and research is important but insufficient by itself. Graduation from an accredited program with a formal degree or certificate in evaluation is not a common phenomenon, given the recent emergence of the profession. Consequently, evaluation experience, professional membership in organizations, and productivity in terms of publications, presented papers, and technical reports have also been selected as indicative of evaluation capability.

Information relating to institutional evaluation resources will also be presented. This includes such factors as evaluation funding, number of professional staff, administrative support, and agency hiring policies. It should be noted, however, that we have not had the resources to fully characterize the professional capabilities of evaluators in federal, state, and local education agencies over an extended period of time and in all their complexity.

4.4 THE ORGANIZATIONAL FACTORS AFFECTING EVALUATION ACTIVITIES AND CAPABILITIES

Before specifically examining capabilities, it is necessary to devote some attention to the organizations which support and conduct the process. The types of evaluation activities undertaken and the manner in which they are performed are at least partly determined by the organizational resources which are allocated to evaluation. The salient aspects include the amount of money committed to evaluation, the number of professional staff assigned to evaluation efforts, where evaluation is placed within the organizational hierarchy, and how this location affects the relationship of evaluation to the program under scrutiny. These interrelated factors play a role in influencing the types of capabilities available to carry out evaluation activities and the quality of the resulting research.

The relationships of money to evaluation capabilities and conduct are not obscure. Federal monies constitute the primary mechanism for supporting evaluation activities and evaluators associated with federal education programs. The amounts allocated to these line items often determine the nature of evaluation efforts and the types of individuals which the organization can attract to carry out the tasks. When individual programs are small in size and funding, evaluation monies may be miniscule. For example, one state official blamed the minimal amounts of evaluation monies (averaging $500-600) as partly responsible for the poor quality of contractors and evaluation reports in the small local Title VII programs within the state. Some states and districts do, however, augment federal funding for evaluation and employ sufficient numbers of highly trained personnel to conduct these efforts. In Site E, evaluation activities for the handicapped programs funded by P. L. 94-142 were improved due to additional district allocations for the evaluation. These supplementary funds permitted the program to hire two full-time professionals from the evaluation unit to work on Special Education program evaluations exclusively.

In discussing evaluation capabilities, it is also important to focus on the location(s) of evaluation within the organization. Where evaluation is placed partly reflects an initial commitment by the administration to producing accurate information and an understanding of the resources required for evaluation. Various types of arrangements exist, placing evaluation "inside" or "outside" the agency. These also incorporate different relationships between evaluation and the program under study. Certain types of arrangements are more prevalent within a particular educational jurisdiction. While evaluations at the federal level usually involve direct grants and the outside contractor model, some efforts are conducted by agency personnel. At the state and local levels, where both state administered and direct grants are employed, there may be a number of organizational arrangements, as outlined by the following diagram.

		Relationship to Program	
		Within the Program	Outside the Program
Relationship to Agency	Agency Personnel	1. Program director 2. Program staff with other responsibilities besides evaluation 3. Program staff whose sole responsibility is evaluation	1. Director and/or staff within an evaluation, research, or testing unit
	Non-agency Personnel	1. Review teams from the federal, state, or county program offices	1. Contractor outside the sponsoring agency

It should be noted that for any given organization, multiple arrangements can coexist either across or within programs. For example, the GAO report of federal educational evaluations found that 27% of the local district Title I evaluations and 31% of the Title VII evaluations were conducted by both outside consultants and inhouse agency staff. Other educational jurisdictions may also play a role in determining the organizational placement for evaluation. In State II, an area populated with small districts, the SEA "highly recommends" that local Title I evaluations be conducted by an outside contractor. Debates are occurring in Congress over potentially restricting the use of external "consultants" and shifting their activities, including the conduct of evaluation, to federal agency staff. The decision concerning where evaluation responsibilities should be located partly centers around issues related to inadequate inhouse resources (e.g., time and staff expertise) and partly around the desire to increase the "objectivity" of the evaluation. We have found that the "inside vs. outside the agency" distinction is less crucial to the independence of the evaluation than the relationship of evaluation to the program under study.

Federal Education Agencies

Federal agencies charged with the support and conduct of evaluations have been enumerated in the introductory chapter. Given the number of organizations involved, our primary focus has been on the Office of Evaluation and Dissemination (OED) because of its major responsibility for sponsoring and conducting evaluations of federal educational programs.

The Office of Evaluation and Dissemination (OED). Prior to the creation of the new Department of Education, the evaluation unit within OED was divided into three subunits: Elementary and Secondary Education, Post-Secondary Education, and Occupational, Handicapped, and Development Programs. Approximately 40 full-time professional staff were employed. The strategy used for carrying out its assigned evaluation responsibilities has been the outside contractor model. This typically involves the competitive bid process with an OED staff member assigned as project monitor. Not only are specific evaluation studies conducted in this manner, but also such activities as the development of state and local evaluation models and the supervision of Title I technical assistance provided to SEAs. It should also be noted that individual projects have been undertaken by OED staff (e.g., the development of a series of technical reports on the packaging of student financial assistance).

Since 1973, OED has administered an average of $21.4 million annually in evaluation contracts. As indicated in Table 1, this money is derived from a variety of sources such as Title I funds and OED's own discretionary funds (labeled Planning and Evaluation). In total, the amounts allocated to the agency in 1979 and 1980 are lower than the peak funding year of 1978. What is interesting to note is the declining monies listed under Planning and Evaluation. According to conversations with OED officials, these funds are typically employed to initiate and conduct studies in response to program managers' requests.

Under the new Department of Education, the evaluation staff formerly within OED are now assigned to the Division of Program Evaluation--one of the three units within the Office of the Deputy Assistant Secretary for Evaluation and Program Management. The subdivisions within this new unit and their overall responsibilities have essentially remained the same. This unit is administratively independent from the various program offices, reporting to the Deputy Assistant Secretary for Evaluation and Program Management.

State Education Agencies

As outlined in the previous chapter, state education agencies play a major role in the evaluation process. Depending on the type of granting strategy employed by a specific federal program, the SEA fulfills such responsibilities as monitoring the compliance of its districts with federal evaluation guidelines, aggregating, analyzing and reporting data on the state-wide impact of federal programs, and ensuring that LEAs receive proper technical assistance in program development and evaluation efforts. The SEA can also engage in its own discretionary evaluation activities for federal programs, and, in many cases, be instrumental in "setting the tone" for evaluation in its respective LEAs. As Mathis and Walling have noted in their report on SEA research and evaluation organizations, the role of the SEA has recently become more proactive in nature. This is at least partly due to such factors as the impetus of court decisions on equal opportunity and funding, public disenchantment with education, and increasing interest in accountability.

Illustrative case studies. Responsibilities for evaluation may be

Table I

AMOUNTS OBLIGATED FOR EVALUATION CONTRACTS OF OED
(In millions of dollars by fiscal year)

Authority	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	(Est.) 1980
Planning and Evaluation	9.7	12.4	12.5	10.5	5.5	6.9	6.3	7.1	7.1	5.2	3.0
Follow Through	4.0	6.2	2.8	3.9	3.1	2.4	2.3	.2	2.0	.5**	-
Emergency School Assist.	-	1.3	.9	3.4	2.5	2.3	1.7	2.1	2.0	1.2**	1.0
Bilingual Research	-	-	-	-	-	-	-	-	.1	1.3	1.5
Labor Dept.	-	.2	-	.2	-	-	-	-	-	-	-
S&E	-	-	-	.6	-	.5	-	-	.9	-	-
Sec. 232	-	-	-	-	.5	.5	-	-	-	-	-
Title I (Sec. 183 & 417)	-	-	-	-	-	4.0	4.5	5.0	12.3*	12.3	15.2
Career Education	-	-	-	-	-	.3	-	.2	.2	.1	.1
National Diffusion	-	-	-	-	-	1.2	.9	1.6	2.1	.4	.7
BEOG's	-	-	-	-	-	.2	.3	-	.3		-
Right-to-Read	-	-	-	-	-	-	.2	.3	.3	.3	.2
Community Education	-	-	-	-	-	-	-	.1	-	.1	.1
Vocational Education	-	-	-	-	-	-	-	1.4	-	-	-
Teacher Corps	-	-	-	-	-	-	-	-	2.4	-	-
Indian Education	-	-	-	-	-	-	-	-	-	.1	.3
Women's Educational Equity	-	-	-	-	-	-	-	-	-	-	.1
Alcohol and Drug Abuse	-	-	-	-	-	-	-	-	-	.1	.1
	13.7	20.1	16.2	18.9	11.6	18.2	16.2	18.0	29.7*	21.6**	22.3

Note: Individual items may not add to totals because of rounding.

* less .55 million transferred to the National Institute of Education for completion of the study of Compensatory Education mandated by P.L. 93-380.

** reduced level in response to DHEW ceiling on consultant services

Source: OED. Annual Evaluation Report on Programs Administered by the U.S. Office of Education: Fiscal Year 1979.

assigned to various offices within the SEA. In some states there is a unit distinct from program divisions while in others evaluation efforts are scattered throughout the agency and nested within the individual programs. The use of multiple strategies is also common, such as program staff for evaluation reporting to federal agencies and outside contractors for supplementary full-fledged evaluations. Information obtained from our site visits may better highlight the diversity which exists.

State II. This SEA serves almost 300 districts and received over $100 million in federal educational monies (approximately 4% of its total revenues) in 1979. All four programs of interest to this study (Title I, Title VII, Special Education, and Vocational Education) exist, and there are also 3-4 state-financed educational programs within the SEA. Although there is a distinct Bureau of Research and Assessment, it is not assigned evaluation responsibilities and only occasionally provides technical assistance in evaluation to individual programs. There is one full-time evaluation position within the Title I Program Office, and there is a separate Title VII Technical Assistance Project Office within the Bilingual Education Program. For Special Education, a Program Development and Evaluation Bureau is housed inside the program. Occupational Education has its own Bureau of Planning, Research, and Evaluation. This organizational arrangement makes the separation of evaluation from the programs under scrutiny impossible. Although many individuals participate in evaluation activities, only three are considered to be "evaluation" professionals by the staff within the SEA. Outside contractors are occasionally employed to conduct special projects for Occupational Education, Special Education, and Title VII. While evaluation is an interest of the central SEA administration, sparking an inhouse study of the agency's evaluation duties and capabilities, there is little move to centralize evaluation within the agency.

State VI. The budget for the Research and Evaluation Office in this SEA totalled over $4 million in 1979, representing an aggregation of 20 separate state and federal funding sources. In federal education monies alone, the state received over $600 million. There are Title I, Title VII, Special Education, and Vocational Educational programs in this SEA. The Evaluation Office has responsibility for evaluating all programs, with the exception of Vocational Education. Almost half of the unit's operating funds are supported by federal monies, along with almost three-fifths of the full-time staff positions. Forty-four professionals are assigned to this office, and 32 of these are employed specifically in evaluation capacities. The office is organized into subunits based on particular program areas (e.g., Bilingual Education Evaluation). Outside contractors are primarily employed to conduct studies mandated by the state legislature for its own state-supported programs. The unit enjoys considerable visibility in the organizational hierarchy, reporting directly the the Chief State School Officer, and is administratively independent from the programs. In fact, it was said that the Superintendent has been very instrumental in fostering the development of this unit.

State IV. This SEA, which governs approximately the same number of districts as State VI, received over $300 million in federal educational monies in 1979 and operates some of its own state educational programs. While there is a distinct program evaluation unit, it is small, having only 5 full-time staff positions. Its primary responsibility is evaluation for Title I and state-supported programs, although several other federal programs with evaluation requirements exist within the state. The unit is administratively separate from the program offices but several layers down in the organizational hierarchy. Its budget is small, and political factors are such that increased support for evaluation in the future appears unlikely.

The prevalence of evaluation centralization and coordination. As indicated by the case studies presented above, an SEA may or may not have a distinct evaluation unit. One source of information on the prevalency of such units is an American Educational Research Association (AERA) survey conducted by Mathis and Walling. They found that the majority of states did have identifiable units for research, evaluation, assessment, and development activities. Our phone survey revealed that 75% of the 36 SEAs contacted had distinct evaluation units or other offices (e.g., Federal Programs) which assumed at least some evaluation responsibilities. Only in a minority of SEAs was evaluation totally nested within all individual programs.

Centralization of evaluation, however, varies in degrees. As indicated in the case studies, evaluation units may neither perform required nor supplementary evaluation activities for all programs. For example, no evaluation unit in our phone survey had responsibility for all four major programs chosen by this study. Approximately one third of the units had some evaluation responsibility for three of these programs, 23% had responsibility for two, and another one third had evaluation responsibility for only one program. A small percentage did not evaluate either Title I, Title VII, Vocational Education, or Special Education programs. Most units (77%) had some involvement for Title I evaluation, and half of the units participated in Special Education program evaluation activities, typically those supported by direct grants. Evaluation of Vocational Education programs was the responsibility of units for only one quarter of the SEAs in states with these federally funded programs.

What this may suggest is the underutilization of evaluation units. While the infusion of federal monies and their associated evaluation requirements have facilitated the creation of these offices, this has not been true across all states which must comply with these federal evaluation demands. Even when an evaluation unit has been established, programs with federal evaluation requirements may choose or be unable to utilize its services.

The nature of evaluation activities assigned to specific programs partly influences the location of evaluation within both the organization and the program. For example, the types of data required by the federal government for Special Education programs tends to place evaluation in the hands of program personnel rather than evaluation unit staff (see Chapter 3 for a description of these activities). However, there are SEAs where the evaluation unit does assume responsibility for supplementary efforts associated with these programs. When there is a competently and adequately staffed evaluation unit, it can improve the quality of efforts, facilitate planning, and increase the use of information in decision-making. Centralization of evaluation also helps in assuring that evaluation activities are not redundant and in developing hiring guidelines which tend to reflect a recognition that certain specialized skills are necessary.

The State Board of Education's and Superintendent's administrative philosophy can help in promoting centralization. When accountability is a concern, a need for valid information is perceived, and policy leadership is a goal, there is usually a move to centralize and coordinate evaluation activities under one roof. For example, in State I the Superintendent adopted a vigorous role by publicly announcing the need for high quality evaluations and the subsequent assignment of them to a distinct evaluation unit with sophisticated skills.

The administrative independence of evaluation. The location of evaluation within the organizational hierarchy differs among SEAs. For example, Mathis and Walling found that 32% of the SEA research, assessment, and evaluation units reported directly to the Chief State School Officer (i.e., the Superintendent or the Commissioner), 56% reported to an Assistant Commissioner/Superintendent, and 12% reported to a Bureau or lower level. Certainly, direct access is important in creating opportunities for policy input and also typically assures that evaluation is separate from programmatic constraints. It may be that this latter aspect--administrative independence--is as important as direct access to policy makers. In our phone survey, over 81% of the units with evaluation responsibilities did not report to the same branch as the instructional components, although a smaller proportion reported directly to the Superintendent. In SEAs where there is no distinct unit, this separation from program pressures to present favorable results can be achieved by the use of an outside contractor. The situation where evaluation is conducted by program staff precludes the possibility of administrative independence for evaluation.

Resources for evaluation. Very little is known about the resources for evaluation within SEAs. Previous research has either tended to only look at evaluation units per se or at SEAs as a whole without considering that some have evaluation units and some do not and that the number of programs for which SEAs have evaluation responsibilities differs among states. Despite these problems, information can be derived from these initial efforts.

For the 32 SEAs reporting budgetary data in a survey conducted by Sharp and Frankel at the Bureau of Social Science Research (BSSR), the median level of research, development, dissemination, and evaluation (RDD&E) expenditures was $374,000 in FY 1976-77. Half of the SEAs reported less than 10 full-time professionals engaged in these efforts. Given that evaluation is only one activity within the agency (according to Sharp and Frankel, about 38% of staff time is dedicated to evaluation and policy studies), this suggests much smaller resources for evaluation. It should be noted, however, that 9 SEAs did report heftier sums of more than $1 million for RDD&E activities, and 13 SEAs reported over 20 full-time professionals employed in these efforts. These still constitute a minority.

Mathis and Walling found that the mean expenditure for units reporting to the Assistant Commissioner/Superintendent level was $2.8 million while units reporting to Bureaus and lower levels averaged $342,000 annually. Research units higher in the organizational chain of command also had an average of 38 professionals as compared to 6 professionals for units reporting to offices further down in the hierarchy. The case studies already presented and other observations from our site visits suggest that resources may diminish as evaluation moves further away from the chief decision-maker. In at least three SEAs, the Superintendents played a major role in creating capable evaluation units. Not only did they commit adequate sums of money to support staff and activities, but they also elevated the unit in the hierarchy so as to guarantee independence, emphasized that program evaluations were to be conducted by the unit, and ensured that high-quality professionals staffed these offices. Capabilities for and conduct of both state and federal programs profit from such actions.

Concerning federal evaluation efforts, our phone surveys revealed that half of the SEAs with research and evaluation units had 1-2 full-time professionals responsible for evaluation activities associated with federal programs. One quarter of the SEAs employed 3-5 professionals, 4% had 6-9 individuals, and 13% had 10-19 full-time professional staff assigned to these tasks.

The federal government is implicated in this process in terms of the resources it awards the SEAs for carrying out their required evaluation responsibilities for federal programs. Our phone surveys and site visits indicated that the majority of evaluation professionals in SEAs are supported by federal money. Funds have also been targeted for strengthening states' abilities in educational practices (e.g., Titles IV-C and V), including the support of evaluation unit professionals. The reduction of federal monies in this area could certainly affect the numbers of staff available to conduct evaluation activities for federal programs.

It was also remarked that evaluation prospers only in times of economic stringency when accountability rears its head. This would suggest that plentiful times are ahead for evaluation and its support. At the same time, however, accountability can assume many shades of meaning--in State IV, a campaign to reduce government bureaucracy has included evaluation positions in its target areas for decreased personnel. Several of the SEAs with which we spoke were concerned over impending staff cuts, especially given the amount of requests for evaluation from both SEA and LEA programs. Caulley and Smith's survey showed similar problems being expressed by SEA evaluation unit directors. We do not yet have an "evaluator index," such as there is in surgical manpower research, which can estimate the amount of professionals required for a specific task, depending on the complexity of the activity. Consequently, no explicit criteria exist for determining appropriate staffing levels in terms of both talent and numbers. Before these complaints of inadequate staffing can be preemptorily dismissed, however, we need a better idea of the resources required for the evaluation process.

Local Education Agencies

Within local school districts there are three typical locations, similar to SEAs, for evaluation: (1) evaluation activities are assigned to staff within the individual programs; (2) there exists a distinct unit in the district, complete with its own staff, whose responsibilities include evaluation, typically for multiple programs; and (3) individuals from outside the agency are contracted to conduct evaluations. A fourth possibility exists, i.e., federal, state, regional, or country review teams or centers which examine district programs. However, this activity typically centers around compliance monitoring for adherence to program guidelines.

The prevalence of evaluation centralization. The UCLA study conducted by Lyon and others provides support for the prevalence of an evaluation unit in districts with over 10,000 enrollments. Over half of the districts had evaluation units, with 91% assuming at least some responsibility for evaluating locally funded programs and 76% for evaluating federal/state funded programs.

However, similar to the situation in SEAs, within any given district the existence of a unit does not always imply extensive centralization of evaluation activities. While in Site G, the evaluation unit shouldered responsibility for program evaluation activity in the district, in Site J the evaluation unit simply provided advice on evaluation for interested programs. Declining resources in this district had reduced the evaluation unit staff from 22 to 2 full-time professionals, and consequently, the unit could not longer be responsible for evaluation activities as was its previous practice.

The administrative independence of evaluation. The organizational location of evaluation within the school district reporting structure has important implications for the autonomy of evaluators. In the UCLA study of district evaluation units, only 37% of the respondents' organizational charts demonstrated that the evaluation unit reported directly to the Superintendent. Generally, evaluation offices were more likely to be in one of the typical lines of authority (e.g., Instruction, Administration, or Support Services) rather than in direct line to the Superintendent. We strongly suspect that direct linkage to the Superintendent may not be as crucial as the organizational branch to which evaluation reports--i.e., administrative vs. programmatic/instructional. In our site visit sample while only 2 of the 9 evaluation units reported directly to the Superintendent, five reported to some intermediary unrelated to programs or instruction. This administrative separation of evaluation from the programs is what helps to guarantee independence of the evaluators from program demands. When the evaluation unit reports to the same authority (other than the Superintendent) as the programs, there may be subtle pressures applied to present favorable results and the linkage of job security to positive evaluation findings. Districts which do not have evaluation units and assign evaluation responsibilities to program staff are also hampered in this respect. The use of outside contractors, when there is no independent unit within the LEA, is one method of obtaining administrative independence. This can be quickly eroded, however, if positive results are perceived by the contractor to be a condition for rehiring, and further employment is the contractor's primary concern.

Factors influencing where evaluation is located. The location of evaluation can play an important role in determining its relationship to programs and the degree of centralization within the district. The existence of a unit tends to increase the possibility that evaluators experience freedom from program demands and pressures. The choice of which arrangement(s) are used depends upon a number of factors. It is partly a function of district size. For example, the UCLA study conducted by Lyon and others found that 89% of the metropolitan districts (45,000 or more enrollments) had evaluation offices while 59% of the large districts (25,000-44,999) and only 33% of the medium districts (10,000-24,999) had such units.

Coupled with the size of the district is the tendency for large and metropolitan LEAs to receive more substantial educational support from non-district sources. In many cases, the more money awarded the district by multiple state and federal agencies, the more likely that evaluation requirements exist which must be fulfilled by districts. The collection and subsequent reporting of this data may be greatly facilitated by a coordination of evaluation efforts. In fact, the advent of federal/state evaluation requirements was cited by 72% of the districts in the UCLA sample as providing the impetus for creating evaluation offices. Two thirds of the evalation units in our site visit sample also indicated that this was a major reason for their establishment. Although requirements are not the sole motivation for centralizing evaluation within school districts, they certainly have played a role.

Variations in federal/state funding procedures for evaluation, the existence of evaluation funding "ceilings," and other evaluation requirements also determine the organizational arrangement and location. For example, there is usually some explicit or implicit and idiosyncratic "rule of thumb" used in assigning the percentage of program funds which can be appropriately targeted for evaluation activities. When total program funding is small (e.g., $100,000), 1-1.5% "set-asides" are typically insufficient for conducting evaluation efforts. Unless the district itself pledges financial support, the program may find itself with inadequate funds to hire its own evaluation staff, procure an outside contractor, or even partially contribute to the salary of an evaluation unit staff member. Thus, evaluation simply becomes appended onto the job responsibilities of existing program personnel who can devote less time to its execution.

Explicit or implicit "guidelines" as to who should conduct evaluations also influence the choice of organizational framework. For example, the Title I evaluation office in State II has "urged" its predominantly small LEAs to employ outside contractors for their Title I evaluations. Only for those districts demonstrating that they have strong evaluation capabilities, as evidenced by a distinct evaluation unit, can this "recommendation" be waived, and the hiring of an external "auditor" for the evaluation itself is urged. In our site visits we found that Title VII program personnel overwhelmingly perceived the use of outside contractors to be a federal requirement, although there is no explicit federal requirement for this. Only in Sites F and B did the evaluation unit conduct the evaluation and, in the latter case, this was in conjunction with an outside contractor. This perception, especially in districts with competent and independent evaluation units, led to disgruntlement. Conversations with OBE program staff suggest that districts are encouraged to choose any strategy which would result in a high-quality evaluation, but project officers may differ in their suggestions to LEAs concerning the best strategy to employ.

The district Superintendent and School Board also can be influential. For example, in Site H, a district where program staff performed all evaluation activities, the Superintendent expressed little interest in program evaluation and saw no need to change this situation for the purpose of coordinating efforts or ensuring independence. In Site C, where there was a one-person unit for a large urban school population receiving $5 million of federal funds, the Superintendent was described as too involved with political skirmishes and indictments to devote interest to evaluation. In contrast, in Site B, a district half the size as Site C, the unit enjoyed considerable support, autonomy, and visibility. The evaluation unit had 10 full-time professionals and a budget triple that of Site C. One reason for this centered around the Superintendent's philosophy and the Assistant Superintendent's extensive background and experience in evaluation. The School Board in Site E recently approved the hiring of six additional evaluation professionals with primary responsibility for executing Board-requested evaluations. Thus, the commitment by these policy-making bodies both determines where evaluation may be placed in the organization and the level of available resources for evaluation efforts.

Resources for evaluation in districts with no evaluation units. In general, programs solely relying on their own staff to conduct evaluations have minimal resources to devote to the evaluation process. This may result for a variety of reasons--few federal/state programs with explicit evaluation requirements operate in the district, those programs which do operate involve small numbers of students, or the district central administration places little emphasis on program evaluation activities. One reason frequently expressed by program administrators in our site visit sample concerned the absence of adequate funds for evaluation. In some cases, even if programs could commit more money to the evaluation process, they are presently struggling to maintain past levels of service delivery, given funding cutbacks. Consequently, evaluation tasks simply become appended onto the job responsibilities of program personnel who seldom have research training.

Seldom is an individual with full-time evaluation responsibilities hired within the program unit. For example, in Site D, until recently there had been a full-time Vocational Education evaluator, but declining resources forced the elimination of this position. Only in Title I programs, where more substantial set-asides are permissible, is this staffing arrangement which allows more concentration of effort made possible. However, independence is still absent in these situations.

Resources for evaluation in districts with evaluation units. According to the UCLA study, the typical evaluation unit's budget is less than $90,000, comprising 0.2% of the district's total operating budget. However, there is considerable variation across districts (e.g., for 1977-78 evaluation budgets in this survey ranged from $2,000 to $4,000,000), and even within similar size districts.

The same study also found that, on the average, only 18% of the evaluation unit's budget is supplied by the federal government, with the remainder stemming from the state and the school district itself. Less than one half of the evaluation unit's budget is actually devoted to evaluation activities, and the typical unit dedicates only 20%.

It is not surprising then that the UCLA study found the majority of evaluation offices to be small, e.g., two or fewer employees. There is substantial variation in these staffing levels, however. For example, in our own site visit observations, units were staffed with anywhere from 1 to 22 full-time professionals. Even in districts where enrollments and federal funding were similar, evaluation unit staffing patterns ranged from 2 to 10 full-time professionals.

Although there is no consensus as to what constitutes adequately staffed evaluation units in terms of the number of professionals, Webster and Holley have developed some informative guidelines. The possession of at least 3 full-time professionals (e.g., evaluation managers and specialists) was defined as a minimal requirement for a 10,000 pupil unit, 4 for a 20,000 student unit, and 17 for a 40,000-50,000 pupil district. Employing these criteria, 67% of the units in our site visit sample were severely understaffed, e.g., 2 professionals for a district with 136,000 students. Given that these criteria were developed six years ago and that evaluation requirements and requests have multiplied, these prescribed staffing levels might even be conservative in terms of present demands. It is not surprising then that evaluation units differ with regards to the number of programs for which they have evaluation responsibilities and the types of activities undertaken.

Approximately half of the Directors of Evaluation units in large school districts reported in the Lyon and others' survey that they did not have a budget sufficient to meet "federal, state, and local requests for evaluation activities." Nearly 70% said their unit's personnel resources were not adequate to meet current evaluation demands. Approximately 94% said they had too few staff. What is important here is that evaluation activities do not only focus on federal/state required efforts but involve much more. Consequently, as the number of requests for evaluation increases, additional staffing may be required. Compliance-oriented evaluation activities in terms of meeting federal/state requirements necessitate a certain number of staff. Engaging in additional evaluation activities for federally funded programs may require even more staff, not only in terms of bodies but also in terms of more sophisticated competencies.

Findings from our phone survey also suggest that staffing levels may affect the quality of evaluation. Districts who took pride in the fact that evaluation activities for Title I were minimal had less than one full-time equivalent individual responsible for evaluation or assigned responsibilities to teachers and administrators. It was boasted that an $850,000 Title I program was "evaluated in one week." In contrast, districts who attempted to engage in efforts which went beyond sheer compliance with federal requirements had at least one full-time professional and often 2-3 individuals assigned to Title I evaluation activities. These were typically the districts where program and evaluation funds were substantially larger than the former cases described.

4.5 CAPABILITIES FOR EVALUATION

This section addresses the capabilities of federal, state, and local education agencies to conduct quality evaluations of federal programs. The absence or presence of an identifiable unit provides the primary distinction, although we recognize that there are several cases in which these strategies are not singularly employed across all programs. Given the attention generated in the use of outside contractors, a separate section has been devoted to examining salient issues.

One underlying thrust of this chapter has concerned examining capabilities in conjunction with the tasks required and their scope and level of sophistication. Institutional resources for evaluation have also been included as a key component. Consequently, general overviews of the assigned activities, both those required by the individual agency and those mandated by other public educational jurisdictions (e.g., SEAs and the federal government), have been provided. Capabilities are then described, both in terms of institutional resources and individual competencies where possible.

Capabilities of Federal Education Agencies

Given that the outside contractor model dominates this sector in terms of evaluation conduct, the focus here is on describing the resources existing in and available to federal agencies which initiate, monitor and coordinate this process. The range of activities for which these individuals are responsible is outlined and general characteristics of evaluation staff described in terms of these roles. It should be noted that many of the following observations are based on the existing structure and employed staff in the new Division of Program Evaluation as of May 1980.

The Division of Program Evaluation. Within this newly created unit, there are at present 38 professionals. These individuals are assigned to the "front office" or one of three subdivisions: 8 professionals are in the Division of Postsecondary Programs, 9 are in the Division of Occupational, Handicapped and Developmental Programs, and 16 are in the Division of Elementary and Secondary Programs. In total, there are 34 middle and senior level professionals and 4 junior professionals.

According to internal policy, the Division of Program Evaluation is charged with the conduct of process and impact evaluations. This is primarily accomplished through the contractor model, involving the preparation of RFPs, monitoring of the study, and production of a final summary. The development of models for SEA and LEA evaluations of federal programs, the monitoring of the state reporting systems on receipt and use of federal educational funds, the monitoring of Title I Technical Assistance contracts, and the provision of technical assistance to other Department of Education offices also fall under this unit's responsibilities.

The actual execution of these responsibilities incorporates a number of activities. Interviews with OED staff concerning their role in the contract monitoring process itself may provide a sense of the diverse activities involved and the capabilities required of these individuals.

Preparation and issuance of the RFP. The project monitor is assigned, based on his/her interest and expertise in the specific study area, to a particular contract. The preparation of the work statement may require such activities as meeting with program personnel in other Divisions and Departments, obtaining advice on technical matters from experts in the field, and surveying pertinent research. A draft of a detailed work statement is prepared and reviewed internally and by selected others. Changes are incorporated, and the final work statement is completed. This includes the objectives of the study, its basic design (e.g., sample selection, instruments, and data collection and analysis strategies), the specification of evaluation tasks, the types of personnel required, supporting documents (e.g., the legislation which initiated the study and program information), and the criteria for proposal review. The RFP is then published in the Commerce Daily, and contractors are given a specified number of days to respond.

Selection of the contractor and award of the contract. The project officer then assembles and chairs a panel of experts to review the submitted proposals. The proposals are distributed to panel members, review criteria explained, and meetings convened. Frequently, before the final decision is made, two proposals are selected, and the respective bidders are requested to submit a "best and final offer," responding to specific questions posed by the reviewers and synthesized by the monitor. Once a selection has been made, negotiation must commence with the contractor concerning such issues as the tasks outlined in the proposal and their execution.

Monitoring of the contract. Monitoring a contract involves overseeing the study during all phases. For example, this may include providing assistance in selecting members of advisory committees and attending their meetings. Monitors can also advise on the development of instruments, design of sampling procedures, and the selection of case study sites. They are the ones responsible for ensuring that the instruments designed wend their way through the maze of formal clearance procedures. This may involve detailed documentation and persistent monitoring of the progress status of the submitted forms. Results of pilot test are reviewed, and the project officer may travel to selected test sites. Eliciting the cooperation of sites may be involved, along with attending training sessions for field personnel. The project monitor must be available to provide advice on any questions which surface during the data collection. We have heard Project Directors' reports of an average of 2-3 telephone calls per week to the monitor and as many as 8 calls per day during critical states of the research.

Once analysis has been initiated, the project officer may assist in designing the analysis strategy and targeting areas which look unusual or especially interesting and require additional probing. In some cases, similar types of assistance are provided to the subcontractors.

Throughout the process the project officer must keep abreast of events within the federal government. For example, Congressional hearings may be scheduled where a presentation of preliminary results would be useful, and it is the monitor's role to guarantee that information is prepared and made available to the relevant parties. Another responsibility is to ensure that the results will provide answers to the substantive and policy-related questions which initiated the study and to also incorporate other important information which subsequently emerged during the course of the research. Approval of contractor invoices also begins with the monitor before continuing through the series of signatures required for payment.

During the final stages of the contract the project officer reviews and comments on preliminary drafts of the report.

Upon completion of the contract. The production of a nontechnical, concise project summary to be distributed to relevant parties is the responsibility of the project monitor. It should be noted that this may not be common practice for many agencies who contract out research. In addition, the monitor checks on how the clearance process is progressing for release of the final report and can be involved in dissemination of the report to relevant parties.

Long after contract termination the project officer may still field inquiries related to the study and provide information to other programs and individuals interested in the findings.

Appropriate capabilities. Given this wide range of tasks, it is not surprising that the monitoring process has been likened to "doing a dissertation four times over." Even this may be a conservative description since project officers typically monitor 2-3 projects simultaneously, depending on the scope and complexity of the individual studies. Performing these tasks requires intensity and breadth of experience. Capabilities include expertise in applied social science research (e.g., statistics, design, sampling, and measurement), sensitivity to the problems posed by field research, and interpersonal and policy-related skills.

The staff in this office generally reflect these capabilities. In terms of formal training, almost three-quarters of the middle and senior level professionals have earned their doctorates or are in the process of obtaining them. Over 80% of these degrees are in research-related areas, with one-half of these specifically focusing on educational theory and research (e.g., educational measurement and evaluation, educational research, and sociology of education) and half encompassing such relevant fields as social psychology and economics. This points to the diversity of expertise which is available within the unit to oversee the evaluations and develop policy implications in the areas of legislation, budgeting, and program development and administration.

Given that formal training in evaluation is an insufficient indication of capability, experience in evaluation and applied research is also important. The staff in this unit have acquired this experience in a variety of ways. The Office itself has often served as a training ground and socialization mechanism for these individuals. Over one-third of the professionals have been in the office for 6-9 years and approximately one-fifth for 10 years or more (at a time when evaluation was a new field of inquiry). Staff have also brought to the agency a wealth of experience from other sectors. For example, almost 75% have had previous experience in LEAs, SEAs, and other federal agencies, and over half of these were employed in specific evaluation capacities. Local and state agency experience is common for those individuals associated with the development of evaluation models and monitoring of technical assistance. One-half of the staff have also had teaching, management, and/or research experience in academic settings, and approximately one-fifth have previously conducted research in the private sector.

Professional activity is also common, and a majority of the staff have been active in a number of professional organizations (e.g., American Educational Research Association, National Council on Measurement in Education, American Psychological Association, and the American Statistical Association). Individuals have produced technical reports, chapters in books, journal articles, and papers for professional meetings. A conservative estimate would show that over half of the staff have published since their arrival in the office. Of these, over two-thirds have produced three papers/articles or more. For example, a few individuals have written 8 or more manuscripts in education since the two years they have been in the unit. This rate certainly rivals that of many academics. Much of this can be attributed to the professional commitment and calibre of the staff and the promotion of this activity by OED administration.

Capabilities of State Education Agencies

As noted in the previous section, the organizational locations and resources for evaluation in SEAs are diverse. While some states have distinct units in which evaluation is housed and some even have further subdivisions focusing on various program areas being evaluated, other SEAs have evaluation responsibilities and staff spread across the individual programs. Even within those SEAs having specific units, staffing patterns vary. In some SEAs the unit enjoys considerable visibility in the agency hierarchy, reporting directly to the Chief State School Office. In others it almost requires a magnifying glass to isolate it on the organizational chart.

Previous research which examined the expertise of SEA research and evaluation staff has reinforced this impression of diversity. For example, Mathis found that 30% of the individuals employed in SEA research-related activities had earned their doctorates, 46% their master's degrees, 14% their baccalaureates, and 10% a variety of other credentials. The greatest proportion of these degrees were not in research-oriented specialties--only 9% were reported to be in psychology, 7% in testing, 3% in statistics, and 2% in computers. The most frequently reported category was administration (29%), with other areas ranging from curriculum to geography.

Reanalysis of the BSSR data revealed that the majority of organizations (74%) engaging in research, development, dissemination and evaluation activities within SEAs had at least one-quarter of their staff possessing doctorates, primarily in the field of education. Evidence of research and evaluation training within this area is difficult to decipher, given that the survey categories were relatively broad. "Education" encompassed such varied areas as educational research, early childhood education, and curriculum. In distinguishing between those organizations which stated that evaluation was a primary emphasis vs. those who did not, "evaluation-oriented" SEAs had slightly greater proportions of their professional staff whose primary expertise was in mathematics and statistics than did agencies not citing evaluation as a focus.

While such overviews of the RDD&E "capabilities" in state educational agencies are instructive, the tasks assigned to these agencies must be considered. SEAs are charged with many responsibilities, only one of which is evaluation. Even this area includes evaluations for both state and federal funded programs. Responsibilities delegated to the state for evaluating federal programs include: (1) monitoring to ensure that LEA plans and programs comply with federal guidelines, as interpreted by the state; (2) annual aggregating and reporting of data for these programs on a state-wide basis; and (3) technical assistance to LEAs in their program and evaluation responsibilities. The range within each of these categories can also vary, depending on the specific program. SEAs themselves can adopt the posture of servicing requests by LEAs for technical assistance, or as in State VI, create a separate state-funded technical assistance program. In addition, a number of SEAs have assumed additional evaluation responsibilities, such as the development of standards and validation procedures, conduct of special evaluation studies, SEA and LEA evaluation staff training, and vigorous contract monitoring. Consequently, the types of activities assigned to the SEA must be considered in judging the adequacy of staff capabilities. SEAs who engage in discretionary and sophisticated efforts require evaluation staff with advanced levels of competencies.

Illustration of time-on-task for evaluation activities. To provide the reader with some idea of the time involved in executing evaluation tasks, data has been presented from the evaluation unit in State VI. This information reinforces the notion that allocations of time can vary across activities and program areas being evaluated. It also shows the substantial amount of time required by provision of technical assistance to both LEAs and other branches of the respective SEA. For example, in the Title VII division, staffed by two full-time professionals and one part-time individual, over 44% of the person-days is devoted to providing LEAs with assistance, and 24% of the time is allocated to the conduct of special studies. Preparation of annual reports to the federal government consumes 19% of the person-days in this unit, with the remainder of the time being devoted to consultation on matters related to assessment, staff development, and liaison with other agencies. Time spent on evaluation activities associated with other programs in this SEA can range as follows: generating evaluation questions (3-8%), designing evaluation studies (3-7%), development of instruments and data collection procedures (8-9%), collecting data (10-25%), analyzing and interpreting data (8-19%), writing, presenting, and disseminating reports (6-20%), providing evaluation assistance to LEAs (20-29%), and providing evaluation assistance to other SEA staff (2-20%).

The preceding example is based on a unit which has evaluation responsibility for most federal programs operating in the state and engages in sophisticated evaluation efforts. Staff capabilities reflect a high degree of research and evaluation training and experience. There are both federal monies for evaluation activities and individual state commitment to the evaluation process. No one can expect SEAs with 2 or fewer staff to launch efforts which match in scope and complexity with this SEA. SEAs lacking distinct units for research and evaluation may be prevented from participating in efforts aimed at planning and coordinating evaluation activities across various programs and servicing departmental requests for supplementary evaluation efforts.

Where evaluation is placed within the agency also helps to influence which individuals will be selected to assume evaluation responsibilities. Program personnel are usually hired for their program expertise, not for their evaluation and research backgrounds. Even when there are full-time evaluators within the program, they may end up devoting much of their energy to program implementation, given that they may know more about the process than the program director. On the other hand, evaluation staff for a distinct unit are often hired with their research expertise and training in mind. For example, evaluation job specifications in States I and VI require at least an advanced degree in research-related fields and professional research experience of all applicants.

Results from our phone survey reinforced this notion--for SEAs with distinct evaluation units, staff characteristics in general reflected doctoral level research/evaluation training for at least 68% of the units surveyed. Within 63% of these units, over half of the staff had earned doctorates in such fields as educational research, social psychology, and testing. When evaluation responsibility was spread across the individual programs, it was much more likely that the staff had not acquired formal training in evaluation. On-the-job training and attendance at workshops were more frequently reported in these cases.

Illustrative case studies. No aggregate statistics reflecting the match between SEA evaluation tasks and capabilities can be offered as this would require a separate study in itself. However, relevant case studies from our site visits can offer a picture of the capabilities which exist in certain SEAs. What will be presented below are short descriptions of three state educational agencies which typify the range and types of evaluation activities occurring at the state level. The capabilities of the staff will be illustrated.

State I. When fully staffed, this unit employs 41 individuals, 71% of whom are full-time professionals. It is lodged in a SEA which vigorously promotes evaluation. Much of this commitment can be attributed to the State Superintendent and School Board who have publicly announced that "evaluations worth doing are worth doing well." Consequently, statements have been issued regarding the need for autonomy of evaluation and expert staff.

Official policy has defined the efforts of the evaluation and assessment unit as follows: (1) the planning and development of the statewide assessment system; (2) the description and evaluation of the impact of state and federal programs; and (3) the assessment of educational needs. To accomplish these goals, the unit is divided into two main branches: assessment and evaluation/research.

The evaluation and research unit is further divided into four components. One focuses on data collection, another is responsible for coordination and dissemination, and two are devoted to evaluation per se. The activities associated with evaluation cover a wide range--both reporting to federal agencies and the conduct of special evaluation studies. Evaluators are typically assigned to a given program and develop an explicit "services of agreement" with the program for evaluation. This helps in evaluation planning, coordination of efforts, and specification of products. Additional studies are also conducted. For example, the Special Education program is receiving assistance by evaluation unit staff in examining the impact of its programs, and special studies on the relationship of certain variables to successful Title I programs have been performed.

At present, within the two evaluation subunits, there are 9 professionals, in addition to the branch supervisor and the director of the entire unit. These individuals have earned their doctorates in research-related fields (e.g., testing, evaluation, and educational psychology). Experience in educational research is common, and many professional papers are produced. The unit has taken advantage of employing interested fellows associated with the national Educational Fellowship Program. Technical expertise, background in educational theory and methods, and other skills exist upon which to draw in conducting evaluation activities.

This is partly fostered by the actual hiring process itself. To be eligible for evaluation positions, individuals must have obtained a M.A./M.S. or 30 graduate hours toward the doctorate in educational research or the social sciences. A specified number of courses in statistics, measurement and design are required. Three years of professional experience in educational or other empirical research or a doctorate is the minimum. An oral examination, described as "comprehensive" and "rigorous," is also required, covering both theoretical and methodological issues.

The fact that state positions remain under the jurisdiction of civil service imposes certain requirements--that all applicants must have gone through the civil service application process. Since this occurs only once a year, it results in applicants passing the requirements but being unavailable 10 months later when a vacancy is finally announced. What also occurs is that the appropriate individual is available and known to the agency, but cannot be hired due to his/her failing to have completed the application process.

State II. Evaluation staff in this SEA are scattered throughout the individual programs. Although there does exist an office devoted to research and testing, evaluation has not yet been incorporated under its job assignments. Only four individuals are distinctly associated with evaluation, and they are located in both the specific program offices and within the research and testing unit itself.

Evaluation activities vary, depending on the individual program. While all programs must report certain evaluation data to the federal government, the type of data required varies. Technical assistance to LEAs in evaluation has become a major priority for most programs in the SEA, but the extent to which this has been accomplished is dependent upon the training of the staff within the program. Only in such federal programs as Title I, where there is auxiliary technical assistance provided by the federal government and the SEA evaluator has had some research training, has technical assistance in evaluation been achieved. There is simply insufficient staff to conduct special evaluation studies within the agency, but fortunately, this SEA is located in a large metropolitan area which can attract reputable and experienced contractors. The awarding of a contract for state refinements to Title I (Section 183) has permitted the Title I evaluator to engage in additional research on Title I evaluation.

The majority of staff associated with evaluation have their doctorates in education. One individual is pursuing a doctorate in research and evaluation at a local university. Based on a study conducted by the SEA itself, there is a need for additional staff evaluation training in order to upgrade technical assistance to LEAs in evaluation. Although staffing levels are minimal and research training uncommon, particular individuals have made contributions and are sensitive to certain evaluation issues. These include the need for more rigorous evaluation requirements for Vocational Education programs, the inclusion of statements regarding the uses of required evaluation activities in Title I program improvement, and the specification of alternative models for evaluation in special education. In some of these cases, outside contractors were employed to provide the necessary staff and expertise to successfully complete these projects.

State IV. While this state does have an evaluation unit, it is not staffed anywhere near the level of State VI--an SEA which has approximately the same number of districts and federal programs but seven times the staff in evaluation. The professional staff in the evaluation unit total five, and this is a smaller number than existed in previous years. Although the SEA is growing in terms of educational enrollments, the new political leadership in the state has embarked on a "reduce government bureaucracy" campaign, and the evaluation unit faces additional staff cutbacks.

This SEA is chiefly involved in complying with federal evaluation requirements. No special studies are conducted with regards to federal programs, and technical assistance to LEAs and other SEA departments is infrequently provided. Eighty percent of the work performed by the unit centers around evaluation reporting to federal agencies.

Although many of the staff have previous work histories in educational organizations, no one has advanced training in evaluation and research. For example, one individual was elevated from secretary to statistician after obtaining a baccalaureate in business. Other professionals have been employed in the unit for a number of years. This is not to say that these individuals are "incompetent," but it does suggest that there is little opportunity to engage in many activities which might improve evaluation practices in the state.

What these case studies should tell us. What these brief sketches should highlight are: (1) the SEA can play a variety of roles in the evaluation process associated with federal programs, ranging from compliance to sophisticated evaluation activity; (2) there are SEAs with the capabilities, interest, and administrative support to engage in high quality evaluation efforts; (3) there are also SEAs interested in improving evaluation practices, and, recognizing certain weaknesses in expertise and resources, which carry out these activities through the use of external assistance; and (4) there are SEAs where activities chiefly focus on compliance with federal requirements due to organizational, financial, and other constraints. In our site visits and phone surveys, we have found that the responsibilities assumed by individual SEAs often match the type of resources (staff, money, and expertise) it has at its disposal. States with no distinct evaluation units or with two or fewer professionals assigned to federal evaluation activities tend to expend their energy toward complying with federal requirements. Exceptions to this are those SEAs which govern a small number of districts or which have program staff specifically trained and interested in research and evaluation. These SEAs also know ways to augment their capabilities. For example, available funds are used to commission special studies of interest to outside contractors with the requisite staff and corporate qualifications, and SEA staff carefully monitor the research. On othe other hand, demands by the other SEA divisions and local districts are not as easily serviced, given small numbers of SEA staff and lack of evaluation training, unless such efforts can be provided by federal sponsors (e.g., Title I Technical Assistance Centers).

State education agencies which display high levels of evaluation activity over and beyond federal requirements typically have the resources to commit to the process, and they deploy these resources in profitable ways. For example, State VI, in using Title V money awarded to enhance state capabilities, pooled their funds with other SEAs receiving similar awards to formulate a coherent and comprehensive list of needs and capabilities and exchanged the needed capabilities among states. Graduate students and other interns are employed to fulfill evaluation tasks. In this way students become better grounded in the realities of evaluation, and SEA staff are exposed to new skills and ideas by the students. University professors also tend to become involved in SEA activities through the work of the interns which they supervise and advise.

To be sure, in our site visits complaints were voiced by districts regarding the evaluation capabilities of SEA staff. These expressions of dissatisfaction, however, usually focused not on the quality of the services being currently provided by the SEA but rather on the additional services which the districts felt should be implemented. For example, large competent LEAs want the SEA to conduct special evaluation studies and issue waivers to districts so that they themselves can execute desired research. Title VII staff in local school districts want to receive technical assistance in evaluation on the same level as their Title I counterparts. Programs which employ outside contractors for evaluation want guidance as to the appropriate selection and monitoring practices in these arrangements. It should be noted, however, that the frequency and intensity of these criticisms often paled against those which were leveled at the federal education agencies in terms of their lack of communication and feedback to the districts which provide them with evaluation data.

Capabilities of Local Education Agencies

In many ways the issues related to evaluation capabilities of local school districts mirror those discussed for state education agencies. For example, the value placed upon evaluation by the central decision makers, as evidenced by the location of evaluation within the organizational hierarchy and the financial resources committed to the process, helps to determine the evaluation activities which are undertaken. These, in turn, select the types and levels of capabilities required to perform these activities.

Local school districts and programs must march to the beat of two drummers--both federal and state program and evaluation guidelines. Many times these two are compatible--in some instances they are not. Compliance with these requirements then forms the common denominator across all LEAs in terms of evaluation activities.

Some districts, however, go well beyond these efforts by analyzing additional program and outcome variables, evaluating programs more frequently than is actually mandated, and exposing their evaluation products to validation and review processes. While the SEA can be instrumental in fostering such evaluation attitudes and practices, it is also the case that individual districts can initiate or facilitate this activity and, in certain respects, surpass the SEA in their evaluation efforts. Districts with such track records constitute a small minority. Considering the large number of districts with federal programs and evaluation requirements (e.g., the NCES survey reported that almost 14,000 LEAs receive Title I funds), the majority of effort often does not exceed what is required.

Given this skewed distribution of activities across LEAs, it is necessary to look at capabilities both in terms of activities associated with adequately complying with federal/state evaluation requirements and engaging in activities which go beyond these mandates.

When evaluation responsibilities are assigned to individual programs and their staff. When evaluation is housed under a program's roof and conducted by program staff, evaluation demands usually are limited to compliance with federal/state mandates. Consequently, tasks are concentrated towards the collection and reporting of data in pre-specified ways. There may often be no need to develop data collection instruments as the SEA itself designs and implements data collection procedures (e.g., Vocational Education programs). While in some programs such as Title I, tests must be selected, administered, scored, and analyzed, other programs simply require headcounts and fiscal accounting. Building staff may participate in such activities by administering achievement tests or employer surveys, and the program staff subsequently aggregate the results or simply give the raw data to the SEA. One classic example of how Title I evaluation requirements are met in small LEAs is that of the school principal who scores the tests, matches the scores, aggregates the results, and completes the required SEA forms.

However, these efforts may not totally encompass the range of tasks involved in evaluation. Some program staff do attempt to collect more information than required. One determining factor is the amount of money available for evaluation. Another key factor is the absence of what has been labeled as the "compliance mentality." For example, while the full-time Title I evaluator in Site H simply aggregated test scores, reported them in the proper format, and occasionally explained testing procedures to building staff, his counterpart in Site J spent only 20% of his time on such tasks. The remainder was used to interview teachers and principals about evaluation practices and observe classrooms for recommendations regarding Title I program implementation. The disparity between these two cases seems to center around the attitude towards evaluation. While in the first site it was stated that "the simpler and briefer evaluation was, the better," the evaluator in Site J perceived federally-required tasks as "mundane reporting" and consequently attempted to go beyond what was required by the state. Title VII staff in Site C, while relying on outside contractors to produce the required report, felt additional information to be important for program improvement and generated (although ineptly) some implementation data themselves. In many sites Vocational Education program staff, who often view state required information as less than useful, were involved in more than the simple collection of enrollments and placement rates, spending much of their time visiting classrooms to detect program strengths and weaknesses. It should also be noted that strictly limiting oneself to compliance with federal/state evaluation mandates may not always reflect disinterest in evaluation, but rather a feeling that the existing resources (in terms of staff and money) are insufficient to engage in efforts capable of yielding accurate information.

Capabilities of program staff to conduct these federal/state required evaluation tasks, let alone engage in discretionary efforts, are minimal. Training backgrounds of program administrators and staff typically reflect program specialties, primarily in the field of education (e.g., vocational education and learning disabilities). There is a striking absence of research training, let alone actual courses in evaluation. Since most are certified teachers, some exposure to testing and measurement was acquired during the undergraduate years, but the time lapse between this training and the present is likely to have been substantial and the content insufficient for evaluation tasks. This should not suggest that these individuals are professionally incompetent--many are active members of such organizations as NASDE and subscribe to such educational publications as Education Daily. The lack of evaluation training merely reflects the fact that these individuals were not hired for their evaluation expertise and most likely did not expect to have this included under their job responsibilities.

Technical assistance that is easily accessible to the LEA can enhance capability. For example, Title I programs relying on staff for evaluation have utilized the Technical Assistance Centers (TACs) to assist them in satisfying federal evaluation requirements. As the NCES survey reported by Goor indicated, in 1978-79 the TACs were called upon to provide rudimentary information concerning student selection, test selection, use and interpretation of normal curve equivalent scores, and proper preparation of reports. More recently, assistance in better using evaluation data and expanding evaluation activities has been offered to interested districts. State education agencies can also be instrumental in encouraging the use of the TACs and some have supplemented these services to LEAs who request additional assistance.

However, programs other than Title I may not have such readily accessible evaluation expertise at their disposal. Technical assistance in some of these programs still remains the responsibility of the SEA and concentrates more on guidance in developing and implementing programs. One exception to this was State VI which provides its own evaluation assistance through workshops and individualized services for administrators, program specialists, and others. However, this has tended to focus on evaluation in general or on specific state programs rather than federal programs. Other resources have been used such as Title VII management workshops and state conferences. At the same time, it must be remembered that these workshops can only include so much in 2-3 days. In addition, school calendars may not always coincide with workshop schedules, and small LEAs who may most need the assistance may not have the interest or the resources to attend.

When funds are available to support full-time evaluator positions, staff characteristics tend to more reflect research-oriented training. For example, in our site visits several of the full-time Title I evaluators within the programs had doctorates in specialties involving at least some exposure to research (e.g., science or math education). Consequently, these individuals may be better prepared to fulfill evaluation reporting requirements and perhaps even pursue some small-scale additional efforts, given that outside assistance for specific design and analysis problems is readily available.

Again, what is problemmatic are those instances where high quality technical assistance in evaluation is necessary but not readily accessible. When there are evaluation tasks to complete which require some research skill, but there is (1) no district evaluation unit to contact, (2) no formal assistance mechanism at the SEA or one which is understaffed, (3) no encouragement by the SEA to utilize federal technical assistance resources, and/or (4) no nearby interested university to call upon, program staff are left with little opportunity to competently execute federal evaluation requirements. For example, in Site C:

The SEA itself was understaffed and until recently had not encouraged its districts to do more than comply. In fact, it was perceived that political tensions were consuming much of the efforts of the SEA. Although the district did have a testing and research unit responsible for evaluation, it was severely understaffed, lacked specific evaluation expertise, and had not developed a good working relationship with the programs. There was funding available for two full-time Title I evaluators, but the Board had prevented these vacancies from being filled. The local universities did not include evaluation in their course offerings or degree programs. While the Title I program staff did use the TACs, they could not optimally exploit this expertise, given that the district administration and evaluation unit were fairly uncooperative in helping them recognize where evaluation could fit into program decisions. Title VII staff, on the other hand, had done some reading in evaluation and knew the types of information that evaluation might provide, but could not benefit from such federally provided technical assistance as their Title I counterparts. The two thousand dollars which had been available for Title VII evaluation could not attract a competent outside contractor. The results was that across the board federal/state evaluation requirements were marginally met, and any additional efforts were few and suffered in quality.

Cases like this suggest that if the federal government requires evaluation and wants quality information to be provided, it must support districts in accomplishing these goals.

When evaluation responsibilities are assigned to a research and evaluation unit. Some information is available on the capabilities of those individuals employed by evaluation units. The UCLA study found that most Directors of Research and Evaluation report having earned doctorates (76%), primarily in the area of educational administration. While 86% of these individuals claim to have had completed at least one course in evaluation as part of their formal academic training, the quality of these courses can vary. A minority of 14% acknowledge any specialization in research and statistics.

With regard to the characteristics of professional staff in general, a variable which was not included in the UCLA research, BSSR found that most staff (58%) in small local education agencies (i.e., 10,000-49,999 enrollment) who were involved in research, development, dissemination or evaluation list education as their major area of expertise. Fewer than 10% of the staff of such agencies cite statistics as their primary specialty area. Approximately one-half of these agencies report that 25% or less of their staff have doctorates. In terms of large districts with enrollments of 50,000 or over, a greater percentage of the staff in these agencies have their doctorates. There is also a tendency to have higher percentages of staff within LEAs whose primary expertise is in math and statistics.

Webster and Stufflebeam, in their examination of evaluation competencies in local school districts with evaluation units, found that urban school districts expect a high degree of technical proficiency from their staff. For example, evaluation methodology and experimental design were consistently rated as among the most important competencies. In addition, large urban school district evaluation directors also wanted such skills as instrument development, multivariate inferential statistics, and computer applications in their staff as a whole. Based on job announcements obtained in our site visits, we also found that urban districts who engage in a variety of evaluation efforts asked more of their applicants in terms of formal academic training in evaluation and technical expertise.

The less compliance-oriented a district is, the more a wider range of skills is needed within the evaluation unit, although no one individual must possess all capabilities. The competencies of the staff typically are defined by the tasks which must be performed. In most cases, the level of skills existing among the staff match the level of expertise required by the evaluation efforts, but occasionally there are individuals who are totally unskilled in evaluation. For example, in Site K, the Superintendent allowed the evaluation unit to become a "dumping ground" for "people whom the Superintendent did not know how to handle or felt were too dumb to do anything else." If the capabilities are too disparate, problems result--either in poor quality evaluations due to the lack of appropriate skills or dissatisfaction and eventually staff turnover due to the evaluators being too highly trained for the tasks required. Inadequate hiring policies, lack of professional incentives, and political obstruction of evaluation efforts contribute to this state of affairs.

Matching capabilities to assigned tasks. Throughout this chapter it has been maintained that the capabilities required for evaluation are highly dependent on the evaluation tasks assigned. Simple evaluation reporting to federal/state agencies does not require the same types of competencies as does engaging in discretionary efforts which examine additional evaluation questions. These latter endeavors also range across districts with regards to their scope and complexity--from gathering information on client satisfaction to estimating the impact of the program and its various components. As the sophistication of these activities increases, so does the level of expertise necessary for their successful execution. Table 2 presents information, based on a review of annual final reports, for the Title I evaluation activities occurring in two local school districts. For both sites, the activities required by the SEA are outlined, along with efforts initiated by the districts themselves. The discretionary efforts described for Site I are representative of those performed by most LEAs which go beyond sheer compliance with federal/state reporting requirements. The profile of evaluation activities occurring in Site II reflects those conducted by the small minority of LEAs which engage in more advanced evaluation practices.

As can be seen from Table 2, obtaining evaluation information for SEA reporting does not require advanced technical and research skills. Simple headcounts, calculation of means and ratios, and conversion of raw test scores to standardized scores constitute the predominant mode of analysis. Evaluators in Site II must also send in the raw data on which these figures were derived. Ways to collect, compute, and report the data are explicitly outlined in Title I evaluation and reporting manuals and further explained by TAC and SEA staff. Some sense of record-keeping and data management, along with an appreciation for accurate information, is warranted so as to ensure the accuracy of the data. Rudimentary exposure to correct testing procedures is also necessary, but much of this is explained in the accompanying manuals. It should be noted, however, that data concerning the prevalence of reporting errors suggest that there is much room for enhancement of the basic skills required.

Table 2

Title I Program Evaluation Activities in Two Local School Districts

Purpose	Activities Required by the SEA*	Activities Initiated by the District	
		Site I	Site II
Program Description	-Obtaining counts of the number of students served and the number eligible but not served (I). -Obtaining counts of the number of: participants by grade, supporting service, and race; parent participants in various program activities (I, II). -Obtaining counts of the numbers of staff employed in various capacities (I). -Obtaining counts of the number of students in different types of programs (e.g., pull-out) for Grades 2, 6, and 10 (II). -Obtaining data on the time devoted to instruction (minutes or hours per week) and instructor per student ratio for Grades 2, 6, and 10 (I, II). -Indicating subject area and length of program (I, II).	-Presenting information on the number of students served and the number eligible but not served by individual schools. -Developing a narrative of program characteristics. -Presenting data on the number of instructional staff by individual schools. -Designing and administrating a survey assessing Parent Advisory Council member satisfaction.	-Presenting information on the number of students served and the number eligible but not served for various instructional components. -Developing a narrative of program characteristics. -Presenting information on the number of students served by various demographic characteristics (e.g., sex).

Table 2 (Continued)

Title I Program Evaluation Activities in Two Local School Districts

Purpose	Activities Required by the SEA*	Activities Initiated by the District	
		Site I	Site II
Examination of Inservice Efforts	-Obtaining counts of the number of inservice training sessions held (I, II). -Obtaining counts of the number of participants by instructional capacity (e.g., aide) and assignment to Title I or non Title I programs (I, II).	-Designing a simple inservice training evaluation instrument to discern whether participants "liked" or "disliked" the training sessions. Analysis focused on simple frequency counts for each item in the instrument. -Designing a needs assessment instrument for one inservice training session to determine what staff wanted from the training module. Analysis focused on simple frequency counts for each item included in the questionnaire.	-Designing and conducting an evaluation of the Inservice Component. Methods included the maintenance of a detailed log of inservice offerings, structured participant interviews concerning aspects of the training they received, the administration of an adapted standardized instrument for measuring satisfaction with the training, and the collection of cost data. Analyses compared expected vs. observed use of the inservice center for the district as a whole and for staff in various instructional capacities. Average satisfaction ratings with the inservice training for the district were compared with city findings and national norms derived by the publishers of the standardized instrument. Satisfaction ratings were also compared among individual presenters. Cost breakdowns were provided per day, per participant, and per type of training and inservice activity.
Reading and Math Achievement	-Obtaining counts of the number of students with both pretest and posttest data (I, II). -Indicating testing interval or dates (I, II).	-Designing and administering a survey to Title I and classroom teachers regarding their perceptions of the program. Analysis focused on frequency counts for survey items.	-Comparing achievement test scores for Title I students with the test scores for students in a comparison group (this groups consisted of students who were eligible for Title I but who were not served).

Table 2 (Continued)

Title I Program Evaluation Activities in Two Local School Districts

Purpose	Activities Required by the SEA*	Activities Initiated by the District	
		Site I	Site II
Reading and Math Achievement	-Indicating the pretest and posttest norm dates (standard or non-standard), levels, subtest area, and forms (same for pre- and posttests) used (II). -Presentation of individual student scores in percentiles (II). -Calculating the NCE pretest mean, NCE posttest mean, NCE gain (NCE pretest mean subtracted from posttest mean), and weighted NCE gain (NCE gain x number pre and posttested/number pre and posttested) by grade (I, II). -Calculating same statistics presented above for individual projects in Grades 2, 6, and 10 (I). -Obtaining counts of the number of students who improved in math, reading, work habits and behavior adjustment as assessed by a 3-point standardized rating scale administered by teachers at the pre- and post-testing dates (I).	-Designing and administering a survey to a random sample of 1,500 Title I parents regarding their perceptions of the program's benefits. Frequencies of the "yes" and "no" responses were reported. -Analyzing the achievement test data by grade in terms of the mean pretest score, the mean posttest score, the mean NCE gain, the average amount of time spent in instruction, and the mean NCE gain per month. -Reporting the achievement or lack thereof for performance-based objectives (e.g., 65% of students in Grade 1 will gain at least 6 and not more than 10 NCE's).	These comparisons consisted of examining the mean pretest scores, posttest scores, and NCE gain. Pre- and post-percentile distributions were also plotted. -Comparing achievement test scores for those students given another year of Title I services although they no longer qualified because spring testing scores were above the cutoff criterion with test scores of similar students who were not given another year of service. Comparisons included same strategies described above. -Comparing achievement test scores for students receiving various instructional strategies in Title I programs. Comparisons were also performed for these groups with a "no treatment" group. Results were analyzed in terms of the mean pretest, posttest, and NCE gain scores and percentile gains. An analysis of covariance was also attempted but judged inappropriate due to extreme differences across the various instructional groups for pretest means. -Conducting an ex post facto analysis of achievement for Title I students and those eligible students who were not served. Grade-equivalent means were presented, t-tests were performed, and a multiple regression

Table 2 (Continued)

Title I Program Evaluation Activities in Two Local School Districts

Purpose	Activities Required by the SEA*	Activities Initiated by the District	
		Site I	Site II
Reading and Math Achievement	-Presenting information on Title I programs which do not employ the norm-referenced model (e.g., description of objectives and assessment procedures and judgment of whether or not objectives were achieved) (II).		was used to compare served and nonserved students while controlling for IQ, prescore, and sex. Grade equivalent averages for students were also plotted against national norms and chance averages. Additional regressions were also performed which examined the effect of years of treatment and the treatment sequence on achievement. -Designing and conducting a methodological study to provide baseline data comparing reading pretest scores and gain scores for Title I students and a comparison group. Statistical procedures used to compare pretest scores to gain scores were Pearson product moment correlation coefficients and analysis of variance and Duncan's New Multiple Range Tests. -Comparing the extent of implementation for 5 Title I instructional approaches. Methods employed included classroom observations and structured interviews with teachers. Analyses involved comparisons of the costs per pupil, attendance levels, implementation ratings, use of support services, and achievement gains across the 5 strategies of instruction.

---

*The Roman numerals presented after each activity in this column refer to the SEA requirements for Site I and Site II.

The discretionary activities performed by evaluation staff in Site I require some research background and training. Knowledge of the rudimentary principles of sampling, survey design, and questionnaire development is important. A review of the instruments and evaluation practices conducted by this site does not suggest the presence of highly advanced research competencies. For example, rating scales were often limited to the dichotomous categories of "like" and "dislike," and analyses never went beyond the simple reporting of response counts for questionnaire categories. Concerning achievement data, analyses focused on simple descriptive statistics and did not attempt to incorporate possible comparison groups or additional factors which might interact with the program. The execution of these activities could possibly be improved by increased technical expertise. However, it is debatable whether sophisticated research training at the doctoral level is warranted, given the types of tasks conducted by this evaluation unit. In addition to technical competency, some skill at managing a number of ongoing research efforts is required.

The activities conducted by Site II and its evaluation unit involve relatively advanced technical expertise. Knowledge of both descriptive and inferential statistics and their limitations is mandatory, and training in such methodologies as correlation/regression strategies is required. A background in a variety of research practices--interviewing techniques, observational strategies, instrument development, and evaluation design--is essential. Expertise in management of large data bases and coordination of a number of evaluation efforts is essential, at least for a few of the individuals in the unit. In addition, unlike Site I, this district assumes the responsibility for formulating recommendations in its evaluation reports which may suggest that some background in educational theory and practice is desirable.

Illustrative case studies. Selected case studies are also presented so as to better depict the types of evaluator skills and characteristics in LEA evaluation units and the extent to which these match the tasks required. Factors which can erode existing capabilities are described. In the first two case studies, district enrollments are similar, as are many of the evaluation efforts conducted by the evaluation units. The disparities lie in the unique organizational and administrative support experienced by the district portrayed in the first case study. The third case study, while based on a district with approximately twice the enrollment, is an example of a "compliance-oriented" district which suffers from a number of constraints. It is contrasted with another district of similar size which engages in a variety of sophisticated evaluation activities and which enjoys considerable local visibility and support.

Case 1. In Site B, the research and evaluation office has been in operation for almost a decade and is highly visible within the organization. It is administratively independent from the programs but does not experience fiscal independence. It is supplied with greater than average resources in terms of funding and allocated staff positions. The budget is 0.6% of the total operating budget of the district, and there are 10 full-time professionals in the unit, including the Director, and two part-time positions. Almost half of the unit's budget is derived from federal sources, with one-quarter each being provided by the district and the state. All of the evaluation positions, with the exception of the Director's, are at least partially supported by federal funds. Evaluation is awarded an unusually high level of sponsorship by the Superintendent and his Assistant. In fact, the Assistant Superintendent has advanced evaluation training and substantial experience in the area, having formerly served as the Director of the Evaluation Unit.

The unit's activities include state and city wide testing, evaluation of federal, state, and district programs, provision of evaluation training to district personnel, conduct of special research projects, monitoring of outside contractors when they are utilized, and general technical assistance to program and building staff. In addition to complying with federal evaluation requirements, they perform such discretionary studies as longitudinal analyses of achievement and examination of the problems associated with evaluating programs where children receive funding from multiple sources. The unit is responsible for the required evaluation activities for Titles I, VII, and IV-C and for Special and Vocational Education programs funded through direct grants.

In the unit there are gradations of evaluation positions in terms of responsibility and skills--1 Director, 1 Evaluation Specialist, 5 Evaluation Assistants, and 3 Evaluation Technicians. Although all staff participate in all tasks (i.e., from specification of objectives to formulating recommendations), the extent of their involvement varies, depending on their capabilities. Staff typically work on evaluation activities for a number of different programs.

The capabilities of the office as a whole match the tasks conducted. Given that assistance to programs and reporting consume a major portion of staff time, interpersonal skills are ranked highly, in addition to the ability to write clearly. Specialized technical skills are perceived to be required in only a few staff who share responsibility for special research projects. No more than two staff have such proficiency, and TACs typically provide sophisticated technical expertise for individual projects. General technical skills focusing on testing applications, basic descriptive statistics, and data processing are necessary for all Specialists and Assistants.

Administration and management of research rests with the Director who has earned a doctorate in education. Originally hired as an Evaluation Assistant, the Director has advanced in the organization, acquiring substantial research and evaluation experience in the local school district. Time allocations for this position are: 50% for management of evaluation activities, 20% for conducting evaluations, 15% for such policy-related activities as attending Board meetings to present evaluation findings, and the remainder on program development and general administration. Coordination of evaluation activities rests with the one Evaluation Specialist who devotes 70% of her time to this effort. This person has a doctorate in educational research.

Evaluation Assistants primarily divide their time between conducting evaluations and providing technical assistance. This position requires the ability to design, implement, and complete evaluations, along with managing the team of staff assembled for the project. Knowledge of norm- and criterion-referenced tests, descriptive statistics, and computer applications is required. These individuals all have their master's degrees, primarily in educational research or the social sciences. Technicians are used for the gathering, analysis, and interpretation of data and are required to have some background in or willingness to learn data processing and analysis strategies. These individuals are currently earning baccalaureates in the social sciences.

In this unit there is a high degree of professional activity and staff development. The majority of staff present papers at professional meetings on testing and evaluation issues, and the Director is extremely active in this area. The unit has applied for and received a NIE grant for research related to objective-based testing in educational programs. In addition, it has sponsored a joint M.A. program in educational evaluation with a nearby university. This program was designed to train students for evaluation positions in school districts and SEAs, and many of the staff participated.

There is no problem with recruitment since a teaching credential is not required, and nearby universities with evaluation programs often provide interested applicants. In fact, the only problem mentioned was low salaries which tended to result in competent individuals being lured into more lucrative positions in other sectors.

Case 2. The evaluation unit in Site C was established in 1973. One half of its resources are devoted to evaluation activities, with three-quarters of this being targeted at evaluations of federal programs. Its responsibilities include state and city wide testing and evaluation of federal and state programs (Title I, Vocational Education, and Special Education). While the unit does report directly to the Superintendent and receives some district funding, it does not enjoy the level of support from the top administration as that experienced by Site B. Evaluation is an interest of the Superintendent only inasmuch as it is one requirement attached to the receipt of federal funds. This lack of interest results in the occasional undermining of the unit's authority and credibility and restriction of the range of activities which can be undertaken.

In addition to complying with federal/state evaluation requirements, other efforts are conducted for the purpose of improving programs. For example, objective-based evaluation is performed, involving goal specification, monitoring of the program, data collection, and assessment to ascertain achievement of goals. Supplementary research on such issues as matching teacher and learning styles has been conducted.

Including the Director, there are 6 full-time professionals, with 3 of these being directly involved in evaluating federal programs in the district. These three individuals have earned their doctorates--two in measurement and evaluation and one in educational administration. The research-oriented degrees reflect extensive training in research and statistical skills. The majority of the staff have also had experience in research and evaluation. In fact, technical and evaluation skills was one criterion for hiring the Director who was recruited at a national conference. The Director is responsible for managing the evaluation efforts in the district and conducting evaluation studies. One of the evaluation staff is hired exclusively for Title I evaluation, and the other individual is assigned to evaluation activities associated with Special and Vocational Education programs.

The problems in this unit do not stem from a lack of training and skills. If anything, the technical sophistication of the staff is underutilized. Activity in professional organizations is common, and some of the staff have presented papers at conventions. It is the organizational constraints which prevent the evaluators from fully exercising their skills. Unlike Site B, where fiscal dependence of evaluation upon programs allows the flexibility for obtaining additional funds for supplementary efforts which improve the quality of evaluation, in Site G such fiscal dependence is problemmatic. The primary reason for the difference between these two sites involves the personalities of the individuals who hold the pursestrings. While the Director of Federal Programs in Site B actively promotes evaluation, his counterpart in Site G seeks to undermine the process. What has resulted is the inability to conduct desired evaluation activities for federal programs and even the loss of one highly trained evaluator. For example, the evaluation unit has been discouraged from hiring an external auditor to review their Title I evaluation report. The Director would like to research aspects of the Title I evaluation models in the district but can not pry loose the funds to do so from the Director of Federal Programs. In addition, staff development is inhibited as expenditures for attendance at professional conferences and TAC workshops must be first approved by the pursestring holder. These problems are not facilitated by the apathy of the Superintendent and have played a role in the resignation of the Title I evaluator--an individual with a doctorate in evaluation. The Director of Research and Evaluation has decided not to fill this vacancy with such a highly trained individual as it will only result in the underutilization of evaluation skills.

Case 3. In this district, almost twice the size of those described in the first two case studies, the evaluation unit was established in the early 1970's. It has a budget which is approximately 0.1% of the total district's operating budget. It does not report directly to the Superintendent, except on an "informal" basis, and is not administratively independent of the programs. It receives minimal support from the policy-making bodies in the district who are constantly embroiled in political squabbles.

The primary efforts of this one-person unit are devoted to system-wide testing and federal program evaluation. Although there are a variety of federal programs operating in the district, the Director is only responsible for testing and Title I evaluation reporting. Technical assistance is erratically provided to program staff, and essentially activities center around compliance.

The possible reasons for this are numerous. First, the unit is drastically understaffed, given the district enrollment and the number and size of federal programs with evaluation requirements operating in the district. The probability of acquiring staff for the two Title I funded evaluator vacancies is slim. In fact, it was stated that it is easier to "obtain new computer equipment than additional evaluators." Much of this is due to local politics, and the Superintendent and the Board have malingered over approving the staffing of even one of these positions.

Second, the background of the Director is in the areas of business and computer processing. Consequently, more of his time is devoted to activities which coincide with these interests--testing and computer programming to score and analyze tests--rather than to evaluation. Additional staff with evaluation skills would certainly be beneficial as they would relieve the Director from the evaluation responsibilities which he does not enjoy.

However, it is difficult to find qualified applicants even if positions were available. Salaries are low, teacher credentialling is required, and former employment in the district is an implicit qualification. Once again, politics only exacerbate the situation by subtly persuading hiring committees to endorse "favorite sons" of the Superintendent who do not necessarily have evaluation qualifications.

Case 4. In Site E the evaluation component is administratively distinct and reports to the Deputy of Management. It has a hefty budget--almost 0.7% of the total district's operating budget. Its responsibilities include the administration of all systemwide testing (located within a subunit of this division) and the evaluation of state, district, and federal programs. The evaluation unit has primary responsibility for all evaluations connected with federal programs, with the exception of Vocational Education. It enjoys considerable administrative support. In fact, the Board recently created six additional full-time evaluation positions to specifically conduct Board-requested evaluations.

The evaluation unit itself views its role as "technical assistance" in the broadest sense. Evaluators help draft objectives, train teachers and other staff in evaluation, explain test scores, and design evaluation components. A number of special studies have been conducted which focus on such areas as Title I evaluation procedures and evaluating specific program components.

At present, there are 14 full-time and 6 part-time evaluation professionals in the unit. In addition to the Director, there are 8 Divisional Assistants and 6 Administrative Assistants I and II. Divisional Assistants are responsible for designing, implementing, and producing the final evaluation product and require at least a master's degree with related work experience and extensive course work in research and evaluation. For Administrative Assistants, requirements are a master's degree or related experience and evidence of a quantitative aptitude. These individuals help in design and analysis and collect data.

There is a broad spectrum of expertise within the unit. More than half of the Divisional Assistants have their doctorates, and some are presently candidates for this degree. Two of the staff have their doctorates in methodological areas (testing and statistics). There are also an anthropologist and child psychologist within the unit. Previous work experience has ranged from evaluation research in a private firm to staffing duties in a state legislature. Former teachers and principals primarily occupy Administrative Assistant positions. The unit has utilized doctoral candidates from nearby universities, and these are usually individuals who fill the part-time positions. The unit is exploring the possibility of establishing a formal internship program in evaluation with a local university.

Staff development activities are common. Once a week staff meetings are held where individuals report on their current activities, present reports of conferences attended, and distribute relevant papers and articles.

Implications of these case studies. These case studies should highlight some salient factors. First, capabilities must be judged within the context of the evaluation tasks assigned and the resources committed to the process. Simple compliance with federal/state requirements, if technical assistance is available, does not necessitate extensive evaluation training. It is when districts attempt to go beyond federal requirements that additional skills are required, and the level of expertise is related to the level of sophistication defined by the evaluation. However, this expertise should be viewed in terms of the unit as a whole rather than each and every staff member.

Secondly, the existence of an evaluation unit helps to ensure that at least some coordination of evaluation efforts occurs, that they are conducted more efficiently, and that some level of technical assistance can be made available to programs. It also tends to recruit, attract, and select individuals with at least some research training. In capable evaluation units the picture is one of "goodness of fit" between capabilities of the unit as a whole and the evaluation tasks. There usually is an adequate amount of technical competency, with one or two individuals being trained in evaluation design, method, and statistics. Outside expertise such as consultants and TACs are used to help solve specific methodological problems. Other staff usually have the necessary abilities to assist in carrying out research activities. Many of the staff may have served as former teachers, a not too surprising occurrence, given that certification can be an explicit or implicit hiring requirement. This teaching experience was often viewed as beneficial by the staff interviewed, especially for providing technical assistance. It helped in not only understanding the program but in being better received by program staff and teachers.

However, what is important to recognize are the various ways in which this match can be weakened. Adequate resources must be channeled to the evaluation unit to hire staff and support some research costs. In districts which have additional resources, the staff tends to be composed of a greater proportion of individuals trained in fields related to education research and evaluation. For example, in Site E (which had a hefty 0.7%--$750,000-of the total district budget), there were 14 full-time professionals, over half with doctorates in educational and social science research. In addition, compare this site to Site C which had the same number of students and types of federal programs but only one evaluation position.

Coupled with this is the issue that many of these professional positions are financed by federal funds. The comment was often made that, regardless of administrative support, if the federal programs and their evaluation requirements were eliminated, districts could not evaluate these programs or could only do so at a reduced level. One reason for this is the financial problems suffered in several LEAs at this time. Consequently, the advent of evaluation requirements has provided monies not only to hire evaluators who, if nothing more, have to respond to federal requirements for information but also to hire evaluators who are highly trained and engaging in many worthwhile and high-quality efforts.

Organizational problems can prevent districts from improving their evaluation practices of federal programs. For example, in some cases the inability to go outside the district to attract specific expertise and/or to hire individuals without teaching credentials limits the capabilities of the evaluation unit and may equip it with individuals who have few evaluation skills and little interest in evaluation as a profession. However, in the majority of districts visited, this was not the most important problem. In fact, in some districts credentialling as a requirement prevented evaluator's salaries from plummeting to those offered by civil service. Other districts circumvent these possible restrictions, as did Site B, by creating evaluator positions which do not fall under the requirements typical for administrative positions.

Given the nature of required evaluation responsibilities, highly advanced training, such as a doctorate in evaluation or methodology, is not necessary for competent execution of these tasks. At the same time, however, it should not serve as a deterrent to engaging in local district evaluation and is required for many discretionary evaluation efforts. In our site visits, complaints were voiced by research-trained doctorates that often their technical skills were underutilized. Some were considering their position as temporary "until something better came along" which allowed greater latitude and creativity. Professional freedom and incentives for evaluation activity need to be instituted so as to prevent units from losing competent staff and to encourage trained individuals to participate in activities which will improve the nature of local evaluation practices.

At present, there exist few avenues for district evaluators who are capable and eager to conduct evaluation research and improve the state of the art and educational programs. As Webster and Stufflebeam have noted, federal funds have generally not proved beneficial in assisting local school districts in answering questions beyond the ones generated from required efforts.

> "The average evaluation department doesn't have the time to play the funding game when, perhaps one in ten proposals are funded. School districts cannot be expected to even write a proposal if the chances of funding are not at least seven in ten. Thus, the dillema continues. Basic research funds continue to be channeled to universities and research and development agencies and not brought to bear on crucial problems in environments where the importance of the dilemma is fully understood and experienced daily."

The ability of LEAs to obtain grants through the competitive grant process as it presently exists was demonstrated by Site B. However, this proposal was developed on off-duty time, and it cannot be expected that all evaluation units can respond in this way. What this does indicate is that there are LEA evaluation units which can design and conduct quality research. These units should be given opportunities to do so.

Universities should also devote some thought to developing mutual programs with state and local education agencies. These programs can improve the capabilities in education agencies and also provide a training ground for graduate students desiring to enter the profession. One comment made by an evaluator with a doctorate in evaluation research concerned the fact that graduate training did not prepare her to deal with the bureaucracy of local schools. Although she did complete an internship while in the program, devoting one day a week for one quarter is insufficient. The establishment of these university and education agency relationships might be fostered by professional organizations (see Prentiss, 1980) or by federal agencies who award training grants or fund workshops to develop evaluation skills. These sustaining relationships may prove more beneficial than the short exposures typically provided by 2-3 day workshops.

The concept of an "endowed chair" or "postdoctoral position" might also be employed in state and local education agencies. For example, one of the staff on this project took a leave of absence from the univeristy to become the Director of Research and Evaluation for a large urban school district. In this way, university professors not only obtain an opportunity to apply their skills, but the agency also benefits from their expertise. Even after an individual's departure, the procedures and ideas remain. These types of positions may be able to provide skilled expertise when hiring restrictions prevent districts from recruiting applicants for permanent positions from outside the district or who lack teaching credentials. Funds might be provided to assist in contributing to the salaries of these individuals if district salary levels are insufficient.

4.6 THE CAPABILITIES OF OUTSIDE CONTRACTORS

Given that the use of outside contractors for evaluation and research activities has generated debate, this section specifically focuses on the contractual arrangement. However, it must be made clear that time constraints have limited our ability to comprehensively explore salient issues. At this time, we can only present some general observations, based on our site visits and interviews, and indicate possible areas which may warrant further consideration.

The Capabilities of Outside Contractors Employed by Federal Agencies

Problems associated with the use of outside contractors by this sector do not appear to primarily concern inadequate capabilities of contractors and monitors. Based on an examination of contracts and interviews with these individuals, we reiterate the conclusion reached by Berryman and Glennan in their analysis of federal educational evaluations. In short, "...few if any unsatisfactory evaluations of federal education programs can be attributed to incompetence or bias of those who fund or conduct them."

It is fairly obvious that the evaluation contracts awarded by OED and other educational agencies have differed in terms of the program area being evaluated, the scope of the effort, the questions to be addressed, and the tasks specified. Often overlooked in discussions of contractual practices are the range of skills and amount of effort needed to mount and execute many of these studies. One major contractor has outlined the required skills as: (1) evaluation and research design; (2) sampling; (3) instrument design, including the ability to deal with the federal clearance process; (4) relations with state and local education staff and school personnel; (5) field operations; (6) data processing and computer applications; (7) statistical analysis; (8) writing and editing; and (9) management of research. These capabilities do not have to be possessed by each staff member, but they must be reflected in the evaluation team as a whole to ensure high-quality evaluation practices. Subcontractors may also be selected to execute specific tasks such as secondary analyses, and the use of advisory panels with expertise in both substantive and methodological areas is a standard practice.

A review of four major studies and six interviews with selected contractors leads to the following generalizations concerning the characteristics of contract research teams and how skills are matched to the required tasks. Project Directors and Principal Investigators--the chief managers and overseers of the research--are typically senior professionals with many years of training and experience in research and management of large-scale studies. They have previously served in a variety of capacities within such institutions as universities, government agencies, and other research organizations. They have published frequently and participated in professional organizations. Middle and junior level professionals, assigned to conduct and supervise such specific evaluation tasks as the development of instruments or analysis strategies, have earned their doctorates in research-related fields and were recruited for their methodological expertise. These individuals have also participated in similar types of research efforts, and many belong to professional societies and publish in their respective fields. Data collectors and research assistants who interview respondents, code and keypunch data, and participate in site visits are usually graduate students, baccalaureates with some quantitative background, or individuals (e.g., former teachers) with experience in local and state education agencies. What this brief sketch should suggest is the manner in which capabilities are matched to the tasks required in an evaluation contract. Competence rests on the goodness of this fit.

It should also be noted how the size of the team fluctuates throughout the course of the contract. During data collection, total staff size reaches its peak. One example given was that of a two-year, $1 million study where the staff swelled from 5-6 during proposal preparation to 70 individuals during data collection and returned to its original number for writing of the final report. Smaller contracts (e.g., $250,000) involve the same numbers during the initial and final stages of the research and approximately 20-30 personnel during data collection. This flexibility is usually accomplished by hiring temporary personnel, and the remainder of activities requiring advanced expertise are performed by full-time permanent professional staff. How research such as this could be accomplished under present civil service requirements and agency staffing levels is unclear.

Problems associated with the contractual process have been noted in these interviews. First, the procurement cycle with its frantic fourth quarter has been blamed by some contractors for affecting both the quality of RFPs and the proposals submitted. One contractor stated that during this final quarter the time to respond was reduced to only 10 days as compared to the typical 4-6 weeks permitted during the rest of the year. Related to this have been complaints regarding the selection and award process. One concerns inequity in the awarding of contracts. However, we have refrained from presenting tabulations which indicate the numbers and sizes of contract awards across various firms. Such aggregate figures can only be misleading, unless background knowledge concerning the types of evaluation activities requested, the resulting professional skills required, the length of the contract, and the quality of all submitted proposals for the contract has been incorporated. Obtaining this information would necessitate a separate study in itself. Problems were also mentioned, especially by contractors in smaller firms, concerning the time from issuance of the RFP to the notification of the award. One example was given where this process took 4-5 months. This can result in situations where staff specified in the proposal have already become involved in other projects.

Delays in contract payment were also cited as problemmatic, especially by smaller firms. One common complaint was that the government and contracts officers sometimes acted as if they were unaware of the contract regulations themselves. Time constraints prevented us from adequately investigating these issues. However, one concern on which we can comment is the need for professional incentives and recognition of quality work produced by contractors. It has been expressed that while the results of contractors' efforts are cited in agency administrators' speeches and in other public addresses, the performers of the research may or may not be mentioned. The OE Annual Report is an admirable example of actually targeting the individuals--both contractors and monitors--responsible for the specific studies. Coupled with this issue is the observation made by many contractors that the periodic uproar over contractor abuse can be very demoralizing to the innocent firms. Contractors felt that the extent of any abuse should be determined and the problems associated with the contractual process which affect the quality of research should be thoroughly examined.

The Capabilities of Outside Contractors Employed by State and Local Education Agencies

The decision to hire an outside contractor usually stems from the need to obtain an "objective" scrutiny and/or specific expertise for the evaluation. However, the existence of an independent evaluation unit with competent staff has increased an agency's own ability to collect data and conduct special studies. Given administrative independence, the need for outside objectivity becomes less clear. This is especially true in light of the possible pressures imposed on outside contractors to produce positive results which will ensure their reemployment for the following year.

In fact, many program and evaluation personnel prefer having the evaluation unit conduct the evaluation. It is felt that evaluation staff better understand the program, are more available to provide advice, and can more quickly remedy any problems which may surface during the course of the evaluation.

Contractors are typically employed by SEAs to conduct special studies of federal and state programs. For example, in State VI, due to an already overworked staff, they have been used to perform evaluations of state programs which were requested by the legislature. Title VII and Vocational Education Programs in State II--an SEA without a distinct evaluation unit--hired contractors to design state-wide evaluation systems which could eventually be used to assess the overall impact of these programs. Less frequent are those situations where contractors are employed to actually aggregate and analyze data for required reporting to federal agencies. The one instance we did find concerned an SEA which had only one district, making the SEA and the LEA essentially one and the same.

In school districts without distinct evaluation units, contractors can be hired to conduct required evaluation efforts--primarily for the collection, analysis, and reporting of Title I evaluation data. For example, the Title I program in Site A has employed outside contractors to obtain the data for Model A reporting. Districts in State II, which typically are too small to have their own evaluation units, used outside contractors for Title I evaluations. In this case, the SEA "recommended" this strategy, suggested that the contractors also be employed to collect process-oriented information in addition to that required by the Title I models, and encouraged that LEAs should hire different contractors every three years to ensure objectivity. Contractors are seldom employed to collect required data for Special and Vocational Education programs due to the type of information which must be reported. When there is an evaluation unit but its primary role is "technical advising" rather than actual evaluation performance, programs hire outside contractors for special small-scale studies. For example, the Special Education Program in Site K used a contractor to conduct a needs assessment of a specific handicapped population.

The one exception to these general practices occurs with regards to Title VII program evaluations. Regardless of accessibility to district evaluation unit staff, local Title VII programs typically hire outside contractors to evaluate their programs and prepare the report for submission to the federal agency. Although there exists no requirement to use outside contractors, district program administrators perceive this to be the case and express disgruntlement over the arbitrariness of this "requirement." Given the reasons previously stated, they see no reason why their evaluation units, especially when they are independent and competently staffed, can not conduct the evaluation.

While we cannot fully delineate the specific capabilities of outside contractors nor derive estimates of the number of "incompetent" contractors, some general observations can be provided. Based on our site visits, review of contract proposals and reports, and interviews, some SEAs and large urban school districts use the same reputable firms as employed by the federal sector. Lists of contractors also include university research firms (e.g., States IV and VI).

Stories of successful experiences in smaller LEAs are less frequent. Instances were given where contractors submitted inadequate two-page reports for a $2,000 contract, refused to rewrite an unintelligible report since payment had already been received, failed to ever submit a report, and used inappropriate methodological and statistical procedures.

There are a number of factors which may lead to the lower incidence of contractor ineptitude we found in SEAs and large LEAs. Contracts issued by SEAs and large urban school districts can often rival the size of a typical federal contract. For example, the average yearly award for 5 current external evaluation contracts in State VI was $163,000, and total awards averaged $552,000. One large urban school district awarded a $115,000 contract for its Title I evaluation. It is not surprising that these contracts were bid on and awarded to firms who are also major contractors in the federal sector.

Even the instances we found when contracts were considerably smaller (e.g., $5,000) did not involve unqualified individuals. One reason for this was that SEAs tend to have better procurement and monitoring procedures. For example, in State VI, the research and evaluation unit devoted 290 person-days to review and selection of contractors and monitoring their progress. In State II which did not have an evaluation unit, individuals in the agency with research and evaluation skills assisted in the preparation of the RFP and selection of the contractor. The bidding and review processes were structured so that the criteria for selection did not focus on cost issues. Payment was spread over the course of the contract and regular progress reports were required.

Problems occur when either contract sizes are extremely small and/or there are inadequate selection and monitoring mechanisms. For example, in many districts, we found the average size of outside contracts to be $2,500, and it was reported that in one state Title VII contracts were usually $500-$600. These levels determine the type of bidders which may respond. Large firms simply do not make a practice of responding to such small-scale contracts, thus, allowing the market to be populated by smaller and lower quality firms. This is not to suggest that small firms are necessarily substandard, but there is a developmental pattern which may exist. As small firms acquire a reputation for quality performance, they are more likely to seek and be rewarded larger-scale contracts which can better support staff and attract competent individuals. In addition, competent firms often recognize that the quality of work suffers when there are not adequate monies devoted to the evaluation.

The following descriptions describe the types of contractors and firms previously employed by the sites we visited:

Firm A. This firm had 6 full-time professionals and specialized in four areas--Bilingual, Title I, Gifted Children, and Special Education evaluations. The firm was created in 1972. On average, it conducts 25 evaluations per year and the typical contract is $4,000. The Director has had experience in state and local evaluations and served as a consultant to some federal agencies. The majority of professionals have degrees in education and have completed some graduate work. However, the most important requirement in staff is perceived to be their previous teaching experience in elementary and secondary schools. Technical skills are not viewed as important since the Director feels that good evaluation resembles "pedagogy" rather than research. He prefers to personally train his staff in evaluation methods and reported that even his secretaries assist in analyzing the data.

Firm B. In contrast, this firm is over 10 years old and only employs one full-time professional. This is because the organization is composed entirely of university professors who do additional consulting. All the individuals teach at one university and have their doctorates. There is one statistician for analysis tasks, a person trained in educational administration to manage the organization and negotiate on contracts, and several curriculum specialists who participate in on-site visits and monitoring activities. The firm never bids on contracts which are under $5,000 or ones in which they are not interested. One reason they can afford to do this is that outside contracting is only a part-time job. Since the staff all have advanced training, however, their fees are considerably higher than those requested by Firm A--$200-250 per day as compared to $45-$85 daily.

The differences between these two firms are not often clearly perceived by selection boards in some school districts. One reason for this is that they often do not incorporate evaluation expertise into their composition. Program directors are typically not trained in evaluation methods which is precisely the reason why they are seeking outside assistance to conduct the evaluation. Consequently, not only selection but also monitoring of contractors can be problematic. School Boards and Superintendents may only complicate the process by inserting political and personal variables into the selection process. For example, in Site A where the Board makes the final decision on whom to hire, the Title I Program Director can only make recommendations and due to district politics feels that his suggestions are often "the kiss of death" for a particular firm. In one site it was stated that the Board tended to glance only at the price tag of the proposal rather than the qualifications of the applicant.

There also exist few sources which local districts and programs can draw upon to assist them in their selection process. Evaluation units can assist in these matters. State II has developed suggested practices which districts should follow in procuring and monitoring outside contractors. Although the SEAs can often provide recommendations, they cannot compose lists of "competent" individuals and delete any who may be questionable. This has resulted in the situation where any individual can call an SEA and ask to be placed on the list of available consultants. These factors strongly suggest the need for the development of standards and guidelines which can be used to inform local districts as to how procurement should be handled, what criteria should be used in judging the quality of proposals, what the district's rights are in the contractual arrangement, and how the monitoring process should be conducted.

Footnotes

[1] Full references to documents cited in this chapter and others are given in Section 8, References, by authors. The text identifies individual authors where possible and the organization that produced the document otherwise.

CHAPTER 5. HOW WELL ARE EVALUATIONS CARRIED OUT?

Robert F. Boruch, David S. Cordray
and Georgine M. Pion

> Just because you can't lay an egg
> doesn't mean you cannot criticize
> an omelette. In fact, being able
> to lay an egg may disqualify you.
>
> Hernando Hibachi

5. HOW WELL ARE EVALUATIONS CARRIED OUT?

This chapter is primarily concerned with the quality of evaluations and factors that influence competent execution, analysis and reporting. In Section 5.1, performance guidelines and standards are described. Section 5.2 reviews the quality of evaluations within the framework of guidelines. Section 5.3 addresses some critical issues, including the need to assess program implementation, the importance of the evaluation design, options for promoting randomized field tests and the need for critique and secondary analysis. The final section discusses constraints on quality of evaluations and suggests options for relaxing them.

5.1 STANDARDS FOR ASSAYING QUALITY OF EVALUATION

Over the last five years, a variety of efforts have been made to develop guidelines on good practice in evaluation. Two of the efforts described here focus on education--the USOE-NIE's Joint Dissemination Review Panel and the Joint Committee on Standards for Educational Evaluation. Two focus on evaluation more generally, covering education, health, law enforcement and other areas. These organized efforts are relatively recent, partly because the whole field is new. But individual university researchers have been working on standards since the 1960's. There have also been earlier efforts, by the Phi Delta Kappa National Study Committee on Evaluation, for instance, to describe the field of evaluation including standards of evidence.

The Various Sources of Guidelines and Standards

The JDRP Ideabook, written by K. C. Tallmadge of RMC Corporation in New Hampshire with substantial advice from NIE and USOE Staff, was issued in 1977 for the USOE-NIE Joint Dissemination Review Panel's use.[1] The manual is a guide for local program developers on the criteria used by JDRP in judging the worth of their innovative programs and in judging the evidence offered in support of a program's effectiveness. Innovations approved by JDRP become eligible for dissemination support. According to Mary Berry, then Assistant Secretary for Education, the guidelines were developed at the request of practitioners.

The Joint Committee on Standards for Educational Evaluation is chaired by Daniel Stufflebeam of Western Michigan University and consists of representatives from twelve professional organizations with interests in standards development. The organizations include the American Educational Research Association, the American Psychological Association and others. Draft standards were developed during 1976-1980 and reviewed by an external panel of experts. The fifth and final draft of the standards will be edited by the Joint Committee and published by a commercial publisher. Contributors expect the standards to be used in college training. Their plans call for continuous revision by a standing committee whose operation will be supported through sale of the monograph. The Committee's work has been supported by grants from the National Institute of Education, the National Science Foundation, and the Lilly Endowment.

The Evaluation Research Society's draft Standards for Program Evaluation were issued in November 1979 by a committee consisting of ERS members, Keith Marvin of the U.S. General Accounting Office is chairperson of the committee which includes university researchers, federal agency staff, and GAO representatives as well as private contractors. The origins of the standards lie in rather strong interests in the topic by some ERS members. The draft has been constructed partly on the basis of other similar documents, notably the JDRP Ideabook, the draft monograph of Joint Committee on Standards for Educational Evaluation and the U. S. General Accounting Office draft guidelines.

The U.S. General Accounting Office released exposure draft guidelines for assessing quality of impact evaluations in 1978. The justification for issuance was the GAO's general mandate to oversee federal program evaluations and the need for clarity in discussing quality. The guidelines were developed by GAO with consultant assistance, and they are dedicated primarily to impact evaluations rather than all types of evaluation. Nonetheless, there is clear topical overlap with other guidelines discussed here.

The American Statistical Association has not issued general standards or guidelines on statistical practice. However, a recent attempt to review quality of a small number of surveys, by Barbara Bailar of the Census Bureau and others, involved development of terse standards which require technical expertise to apply. There is some overlap between GAO guidelines and these criteria. But because the ASA criteria are not advertised as guidelines or standards, we have not considered them here.

Differences and Similarities Among the Guidelines

All the guidelines have certain features in common, topics which the evaluator is encouraged to address. These topics include: description of the program under evaluation, the rationale for choice of evaluation plan and measures, an explicit plan or evaluation design, the inclusion of data on reliability and validity of measurement, full and balanced reporting, linkage between evidence and conclusions, and thoughtful interpretation of results to major audiences for the evaluation. In brief, most guidelines can be classified as bearing on Accuracy, Utility, Propriety, and Feasibility, in the way the Joint Committee on Standards classifies its standards. Some ethical standards are present in each set of guidelines. Attention to individual privacy of respondents and to the public's interest in access to reports and to statistical data for competing analyses, for example, is explicit.

There are differences in detail among the guidelines. The Joint Committee on Standards puts more emphasis on context of the evaluation and practical procedures than others. Brief illustrations of proper adherence to each standard are plentiful in the JDRP Ideabook and the Joint Committee Standards. Fewer examples appear in the GAO document and virtually none appear in the ERS document, but neither of these were financially supported at the level of the other two.

Commentary on and Experience with Guidelines

Each set of guidelines has been issued by its sponsoring group with a request for commentary and criticism. The USOE-NIE JDRP Ideabook was issued earliest and adopted for routine use only in the Joint Dissemination Review Panel operations. Three of the six federal agency staff who are well acquainted with the Joint Dissemination Review Panel and who were interviewed as part of this Project generally say that the guidelines have improved the quality of presentations to the JDRP.

One of the few field applications of such guidelines was undertaken in a recent field survey by Catherine Lyon and others at UCLA's Center for the Study of Evaluation. They condensed guidelines developed by individual academic researchers and by the Joint Committee on Standards for Educational Evaluation to obtain a reduced list of thirteen standards. The reduced list does not differ appreciably from the list, given earlier, of elements common to standards discussed here. It was used to judge the quality of over 100 reports issued by evaluation units in large school districts. More remarkable, the center staff investigated the reliability of judges' ratings based on the standards. After brief training in the use of standards, interrater reliability was found to be high, ranging from .80 to 1.00.

The Joint Committee's standards have been field tested at least informally in 29 sites. Generally field tests involve trying to apply the standards to evaluations or evaluation reports. Four "National Hearings," meetings of professionals, have been held to discuss these standards and to improve them over a two-year period. Commentary from the Hearings and letters containing reactions to the standards are generally encouraging though there are exceptions. Both have been compiled in manuscript form for distribution by Jeri Ridings at Western Michigan.

The ERS standards are undergoing review by interested members of the Society and were reissued in May 1980. The effort is supported by membership rather than foundation funds.

In our field research, we encountered no substantial familiarity with standards or guidelines at the local level. School boards, program directors, and others at the local level are likely to trust the evaluation rather than to examine it relative to formal standards of evidence. Congressional staff members may be less likely to trust evaluations, but even the GAO standards are not especially salient for agency staff members or Congressional staffers. The unfamiliarity may be attributable entirely to the fact that development of guidelines is very recent. Nonetheless, it does seem sensible to make tentative guidelines generally available and to make sure they are understood. This is especially crucial at the federal level to reduce unnecessary argument about what quality means in this area.

Debate Over the Utility of Guidelines

Judging from the experience of the Joint Dissemination Review Panel, the JDRP guidelines are useful in telling program developers what kind of evidence on program effectiveness is credible and what kind is not. The context is special in that adherence to the prescribed standards facilitates favorable review and dissemination of the developer's product.

There is some debate about usefulness of guidelines of standards outside this context. Daniel Stufflebeam, chairman of the Joint Committee on Standards for Educational Evaluation, for example, catalogs the following as criticisms or alleged shortcomings: The standards may promote a field that is not needed and legitimate practices that may be harmful. They may concentrate on minor matters, encourage bad practices which are not explicitly proscribed, and impede innovation. The alleged benefits include making language and definitions clear enough to facilitate communication, and establishing a common frame of reference and acceptable rules for dealing with evaluation problems. Standards may also serve as a basis for monitoring evaluations and to enhance credibility of the process and product.

Similarly divided opinion surfaced in our interviews with federal agency staff and Congressional support staff. A Congressional Budget Office staff member pointed out that guidelines can be constraining: Some evidence, regarded as poor under sensible technical standards, might be entirely adequate for some policy purposes. Moreover, guidelines are a coarse simplification of what we understand about quality of evidence and that simplification may be regarded as sufficient by evaluators who could otherwise do much better work. The major risk, according to a federal agency staff member is that standards can only be useful if there is some agreement on them by competent evaluators.

Our general conclusion is that there is a fair amount of agreement among groups working on guidelines about what should be considered in an evaluation. The guidelines themselves are sufficiently promising to warrant their being field tested and encouraging their use.

They are also sufficiently promising to justify their being explained to interpreters and users of evaluation results. At the local level, this includes program directors, superintendents, school boards, and the like -- if there is sufficient interest. At the federal level, this includes program executives, Congressional support agencies and staff.

It is not clear that guidelines are appropriate for incorporation into law or regulation: Their function is advisory. It is sensible to assure that legislative and regulatory language is consistent with guidelines.

5.2 QUALITY OF EVALUATIONS

The quality of evaluations can be assessed at different levels of effort. These range from overall judgments made by experts, through systematic assessment of the contents of specific reports, to reanalysis of the raw data generated by an evaluation. The expertise and time necessary to perform intensive reanalyses is sufficient to warrant a step-wise approach to quality assessment--that is, successively more stringent levels of review should be considered. If a study "passes" the simple review procedure, it becomes a candidate for more rigorous assessment. The rationale and justification for this approach is simple: serious problems can generally be assayed through an expert review of the report. Some problems are so damaging that more detailed reanalysis is simply not warranted.

Previous Examinations of the Quality of Evaluations

Because evaluation is a rather young enterprise, and because standards have only recently been developed and field-tested, it should not be surprising that comprehensive assessments of quality are rare. Consequently, we have to rely on a variety of sources.

The General Accounting Office. In April, 1975 the GAO surveyed program representatives at local and state education agencies regarding the quality of evaluation reports produced by state and local agencies for Title I, III, and VII. A national statistical sample of 832 local school districts and all SEAs were requested to complete a questionnaire survey. Respondents were those individuals who were familiar with each program. GAO's questions about credibility of the findings referred to the respondent's confidence in the soundness of the methods and reasonableness of conclusions. The questionnaire defined "qualification of findings" as the extent to which the results were properly qualified, assumptions made explicit, and the evaluator described the conditions under which the findings were not applicable.

Table 1 presents the percentage of program officials who rate the two aspects of quality as "adequate or better." In the survey, the local and state program officials were asked to rate the quality of local and state reports, generating cross-level judgments of quality. That is, GAO obtained local views of state reports and state views of local level reports and ratings of quality pertaining to evaluations conducted at the same level of government as the raters.

The GAO findings are interesting on two counts. The judgments pertaining to the quality of the evaluations are consistently higher for same-level reports than for cross-level reports; indicating that judgments on quality may be confounded by the utility of the information at each level of government. Of more relevance to the issue of quality, however, is the fact that even those very global assessments are not complementary with respect to the quality of evaluations. The highest rating of "adequate or better" was ascribed by only 69% of local program officials to local reports on Title III. The lowest level of quality was ascribed by State Officials to local Title I reports; only 31% of the reports were rated as "adequate or better" for the manner in which evaluators qualified their results. The remaining judgments are evenly distributed between these two extremes.

Lyon and others, Center for the Study of Evaluations. Lyon and others requested evaluation reports from each respondent in their survey. They received 116 reports which were then reviewed according to the presence or absence of criteria considered to be necessary elements of an evaluation.

Table 1

Percentage of Program Officials Rating Quality as "Adequate or Better" for State and Local Evaluation Reports: GAO (1977)

Credibility		Program		
	Source of Report	Title I	Title III	Title VII
Local View	Local reports	62%	69%	53%
	State reports	55%	64%	39%
State View	Local reports	33%	40%	-
	State reports	41%	60%	-

Qualification of Findings		Program		
	Source of Report	Title I	Title III	Title VII
Local view	Local reports	59%	66%	61%
	State reports	50%	61%	41%
State view	Local reports	31%	40%	-
	State reports	53%	53%	-

Their summary of the percentage of items appearing in the 116 reports is reproduced as Table 2. The CSE review suggests that simple standards are not often adhered to. Evaluation reports often do not uniformly describe program implementation, address issues of reliability and validity of the data sources, and connect the evidence with conclusions. What appeared to be best about reports is that the great majority "identified program and evaluation participants, data collection sources, data analysis procedures and results." These last three items are uniformly required in the state reporting forms filed by LEAs and labelled "evaluation reports." If these reports were part of the sample reviewed by Lyon and others, we concur that they are quite sparse in that a narrative description of the program, recommendations, and conclusions are rarely present.

Other Estimates of Quality. The approval rate for the projects submitted to the Joint Dissemination Review Panel (JDRP) can be viewed as an upper bound estimate, although biased in the direction of higher quality due to voluntary submission, of the quality of outcome evaluations at the local level. Specifically, 421 submissions to JDRP, 245 (57.7%) have been approved. Discussion with staff members affiliated with JDRP suggest that the quality of evidence brought before the panel, in recent years has improved, nevertheless, the approval rate is substantially below 100%.

A number of attempts to identify exemplary programs also provide partial evidence on the issue of the quality of evaluations. The Vocational Education Equity Study (1979) performed by the American Institutes for Research (AIR) identified, through 250 nominations, 100 "promising" programs. Although their selection criteria was not exclusively based on methodological considerations, 36 programs passed their initial screening criteria. Ultimately, twelve were selected as case studies, "mainly because of the presence of more complete data and documentation" (p. 2, Vol. III, 1979).

Earlier studies by Campeau and others at AIR and Wargo, Campeau and Tallmadge also through AIR, report considerably lower percentages of studies that could be considered methodologically sound. Campeau and others, examined 175 programs in Bilingual Education, funding only 8 programs "judged to merit site visiting." For Wargo and others, the ratio of successful evaluated projects compared to the number that were reviewed was considerably lower.

A recent attempt to isolate exemplary Career Education Activities was undertaken by Hamilton and Mitchell through an OED contract to AIR. Through nominations made by a variety of personnel at federal, state, and local agencies, 394 Career Education projects were identified. Reports were solicited from the directors of nominated activities. A three-phase review procedure produced a list of 20 projects that passed most of AIR's criteria (but not all) ultimately 10 were submitted to JDRP for approval. (11 actually passed all the criteria, but one had already been submitted to JDRP). Seven of the 10 projects were approved by the panel.

Table 2

CHARACTERISTICS OF DISTRICT EVALUATION REPORTS

Percent of reports (n=116) with each characteristic*
Source: Lyon et al. 1978 (CSE, UCLA)

		YES	NO/ CAN'T TELL
1.	The program or product or other object under study in the evaluation is described so that its objectives are clear.	38%	62%
2.	The program or product or other object under study in the evaluation is described so that the form of its actual implementation is clear.	17%	83%
3.	The purposes of the evaluation are described; purposes may be stated in terms of the evaluation questions or objectives.	53%	47%
4.	Audience(s) for the evaluation information are identified.	35%	66%
5.	Participants in the educational program and the evaluation study, and how they were selected for participation, are described.	70%	30%
6.	Data collection sources, such as tests, records, or observation forms, are identified.	92%	8%
7.	The data collection sources are comprehensive enough to answer the evaluation questions.	46%	54%
8.	The reliability of the data collection sources, and the validity of the data collection sources for the purposes intended is described.	10%	90%
9.	Data analysis procedures are described or are evident (as in detailed tables).	81%	19%
10.	Evaluation results are described or presented.	97%	3%
11.	Conclusions or recommendations are drawn from the results.	52%	48%
12.	The congruence of the conclusions with the information provided is described or evident.	28%	72%
13.	The written presentation of whatever was done in the evaluation is clear (even if standards above were not met).	65%	35%

---

* [Percentages may not add up to 100% because of rounding errors.]

We conclude from these assessments that the procedures employed in the majority of evaluations are insufficient for judging the effects of the projects or programs on children. Among the items consistently reported as reasons for rejecting a project from consideration as an exemplary activity were the lack of statistically valid analysis of outcome data and flawed evaluation designs.

A Case Study of Three Contracted Evaluations

In the course of our field visits, studies containing a variety of flaws were uncovered. While extensiveness of the flaws detected in some aspects of this case study is not representative of the reports generated by the research units we visited, it is nevertheless informative as to the types of errors that appear.

The case is based on evaluations of a bilingual education program in an urban school district. Responsibility for evaluation was given to a contractor, selected through an annual competitive bidding process. Each year a different contractor undertook the evaluation. Over the three years there was gradual improvement, attributable perhaps to better contractors, better selection procedures, or more sophistication in the LEA.

The evaluation for 1976-77, done by a Manhattan-based firm, is clearly the worst of the three. The bilingual program or its objectives are not especially well described and the objectives of the evaluation are rhetorical. Descriptions of procedures, sources of information, and the like are weak. The attention to negative aspects of the program is negligible. The technical aspects of the evaluation are inept at best. The following phrases appear in the report. Despite the technical jargon, they are misleading at worst and meaningless, at best.

- "Any increase, whatsoever, in percentile ranks is a significant increase because the percentiles are adjusted per age."
- "School B showed an average increase of 46 percentile ranks while the School A children showed an average increase of only 15 percentile ranks."
- "Anything less than 95% certainty is not considered to be significant growth."
- "Seven youngsters were pre and post tested...the p value of .001 is the highest one can possibly measure with inferential statistics of this nature.
- "Inferential statistics analyzes trends..."

The 1977-78 evaluation is a bit of an improvement. The program and its objectives are depicted in tabular form, the objectives of the evaluation are described in a general way. Achievement tests are described. But the participants in the study aren't identified, reliability of data sources is not mentioned, and specific data analysis procedures are not described. The prose is dreadful. We are told that "evaluation is a formal program...which can then be fed to personnel...," there are lots of references to "meaningful" results, some thanks are offered to the "Project Coordinatoress." The technical common sense is negligible, e.g., "all of the scores ...are statistically significant...," "t-scores were used when the numbers were not large enough to use z scores. A t-score is widely used standard score in which the mean is 50 and the standard deviation is 10."

The contractor was candid in recognizing that some student achievement objectives were not achieved. However, no attempt was made to explain any of these. The increases in achievement were attributed to the program without any recognition of competing explanations.

The 1978-79 evaluation was done by another Manhattan-based firm for about $5000. It involved the following activities:

- Field observation of program classrooms in three periods to determine whether teachers adhered to reasonably sound pedagogical practice in organizing and teaching students.
- Interviews and questionnaire surveys of program staff, school and district administrators were used to determine their satisfaction with the program, character and quality of management.
- Interviews with parents to determine the level of their participation, existence of required advisory groups, and parental opinions. Group evaluation sessions with program staff to obtain some sense of accomplishments and problems.
- Review of program documents to determine if pertinent materials were available; how new pertinent material might be obtained from other districts.
- Analysis of student achievement and assessment of changes in self-concept and knowledge of culture.

This report meets reasonable standards for evaluation in that: program objectives are reasonably clear, the program itself is described, the purposes of the evaluation are described, respondent groups are identified and procedures for collecting information specified. There was some clear attention to validity of the information, results are described, conclusions were drawn and recommendations were made. The written presentation is clear. The contractor's judgments about whether students benefited academically from the program are based on measurement of achievement before entry and after nine months of the program. The procedure used is reasonably clear.

There is no recognition that student performance could have increased in the absence of the program. The conclusions they draw that the program generated increases is not warranted based on their data. They may be correct in stating that the program is a success in improving achievement but the data are not sufficient.

The report was candid in specifying which procedures were adequate, which were not, and in identifying flaws; they were not stressed, however. It was candid in identifying flaws in the program and was fairly explicit about conflicting testimony. The latter includes recognizing parents who agreed that their children's gains were attributable to the program and those who did believe the program was instrumental. The report was candid in reporting vague lines of communication among administrators and staff and the way this might affect operations and student achievement.

Educational Standards for Success vs Comparative Assessments

A common practice in educational evaluation is the establishment of criteria that are to be reached in order to declare the program successful. Gene Glass describes the fallibility of these procedures, especially as they relate to mastery level or minimum competency based evaluation. Because of the pervasiveness of these practices, although usually coupled with other designs, such criteria deserve at least some comment. The basic argument is that setting values such as "80% of the children will read at the third grade level" is simply an insufficient basis for judging the effectiveness of a program. As a basis for judgment, these standards are too arbitrary. A program can be declared successful or unsuccessful simply by setting the criterion at low or high end of the continuum. A few examples that we encountered may help to clarify the difficulties associated with this method of assessing effectivness.

Example 1 "The Title I participating students will exhibit a mean gain of 1 NCE in Math-enumeration on the C.A.T."

Example 2 "As measured pre-post with the Title I Needs Assessment rating scale for pupil status in reading, of the children rated poor or serious by their regular classroom teacher, 52% will be rated at least one level higher."

Example 3 "Preschool children will achieve (on a classwide average) 23 items on a Language Development checklist."

Example 4 "Ninety percent of preschool children, will according to the teacher's judgment, exhibit development toward positive self image and interpersonal relationships."

The use of these types of standards is particularly troublesome in the context of newly proposed, innovative programs where there is little prior knowledge as to what is an acceptable level of performance. Within ongoing programs, for example Title I, the year to year modification of level of performance may successively approximate a reasonable standard, but nevertheless, such standards are insufficient for judging program success. Testing level of competency before and after the program (Examples 1 and 2) is an improvement over the after-only strategy (Examples 3 and 4). But is still insufficient for attributing the gain to the program. Other competing explanations such as normal growth are as plausible in accounting for the gains, as the program. An example of when the criterion based assessment is a more valid basis for judging effectiveness is when a comparison group has also been assessed. That is, a standard (e.g., 80% success) set by an educator, advisory council, and/or through experience can be meaningfully interpreted only when program participants' performance is assessed against the performance attained by those who did not receive the program. However, even with the use of comparison groups, unless constituted through a randomization procedure that is maintained throughout the duration of the study, evidence of success may still be questionable due to other rival interpretations not controlled by the evaluation design.

Impact of Regulations, Requirements and Legislative Mandates on Evaluation

The description of the regulations governing evaluation presented in Chapter 3 identifies issues that have implications beyond the specific programs that were considered.

Once said, it is obvious that the types of evaluation practices that are prescribed will influence the quality of evaluations carried out. It is important to recognize that some of the guidelines contain statements that are inconsistent with good research practices. For example, the Bilingual Basic grants to LEA's regulations specify the use of comparison groups to estimate what performance would have been in the absence of the program. In the next line, "statistical and historical comparisons" are identified as examples of presumably adequate means of deriving such an estimate. These procedures are notoriously subject to statistical biases and other pervasive threats to the validity of the conclusions. If details of this sort are to appear in regulations, deliberate attempts should be made to have them reviewed by methodologists so as to avoid encouraging the use of weak assessment strategies.

For three of the four program regulations we reviewed, data in the form of test scores, head counts, type of service rendered, and so on are to be gathered for the purpose of aggregation. Specific reporting requirements can yield data which can be aggregated to the national level. For the recipients of comparable data (i.e., states and federal agencies), regulations serve a useful purpose. If on the other hand, the purpose of requiring LEAs and SEAs to collect information (and report it) is to stimulate program improvement through the use of evaluation data, minimal reporting requirements (as in Title I) will yield little useful data for that purpose. That is, test scores, by themselves, do not provide useful information as to why gains or losses were observed. On the other hand, for those local agencies who have little interest in evaluation, regulations will serve as a minimum standard for compliance. If the regulations are too demanding, given the available resources, reporting is likely to yield inadequate data or require extensive technical assistance.

Judging from recent published commentary (Barnes and Ginsburg, 1979; Cross, 1979; Linn, 1979, and Wiley 1979) regarding legislative requests for the development and implementation of evaluation models as part of the reporting requirements, there appears to be a need for a formal review mechanism where legislative representatives, those persons devising the models, users, and critics can assess the informational relevance of the product. Further, pilot tests, using a representative sample of sites (not volunteers) should be routinely conducted to assess the feasibility and desirability of employing evaluation models.

The Evaluation Plan and the Proposal Review Process

Direct grants usually require an evaluation component. In most cases, these grants are awarded for the purpose of demonstrating the feasibility of an innovative educational program or for providing special services. Ultimately, those programs that are superior to traditional practices should be adopted by other educational agencies. In order to ensure that the educational value of an innovative program is understood, a well designed evaluation plan should be articulated. This would include, ideally, an evaluability assessment of the program, and the collection of process and outcome information. Some of these are overlooked in practice. Further, the need for careful program planning and evaluation planning prior to the implementation process appears to be undervalued judging from characteristics of the proposal application process for direct grants.

For example, regulations for direct grants to LEAs under Title VII (Bilingual Education) provide a summary of the point values assigned to each review criterion. As was indicated earlier, for basic grants, the evaluation plan was allocated 15 of 110 possible points. Here, the evaluation plan contributes a rather insubstantial amount to the final point-total within the review process. Judging from this case, the selection of projects for funding seems to be more heavily weighted towards the substantive managerial, staffing aspects of the proposal.

To assess the pervasiveness of these practices, we examined the review criteria for additional direct grant programs. These included four programs under the discretionary grants provision in Vocational Education and eight direct grant programs funded under ESEA, Title VII. A summary of the point values ascribed to the evaluation plan and general methodological/evaluation considerations is presented in Table 3. From the entries in Table 3, it is seen that the grant application review process entails assigning between 100 to 110 total points to each proposal, the number of criteria used ranges between 5 and 11 and the points allocated to the sufficiency of the evaluation plan (all but one program explicitly mentions the evaluation plan as a criterion) ranges between 5 and 15. That is, at most, 15% of the review process is devoted to the adequacy of the evaluation plan.

Given the diverse meanings attached to the term evaluation, it may be appropriate to examine the regulations in more detail. If we consider any criterion that contains even the slightest mention of a methodological

Table 3

Weight Assigned to Evaluation Practices in Direct Grant Applications for Vocational Education and Bilingual Education

	Characteristics of the Application Review Process					
	Total points	Minimum points	Number of criteria	Points for the evaluation plan	Number of criterion listing methods	Total points for methods
Vocational Education						
1. Program Improvement	100	50	11	8	2	26
2. Indian Tribes	100	30	9	10	3	35
3. Bilingual Vocational Education Training	100	30	8	12	3	37
4. Bilingual Vocational Instructor Program	100	50	10	10	3	35
Bilingual Education						
1. Basic Grants	110	70*	7	15	2	25
2. Demonstrations	100	NA**	10	7.5	2	12.5
3. State Technical Assistance	100	50	7	15	1	15
4. Support Services Projects	100	NA	7	10	2	25
5. Training Projects	110	NA	7	5	1	5
6. Short Term Training	100	NA	6	5	1	5
7. SEA Training	100	NA	6	10	2	40
8. Schools of Education	100	NA	5	0	0	0

*Established for 1978

**Not available at this time.

issue, we find that, at most, three criteria are contained in the application. This more liberal definition of methodological concern yields a point distribution that ranges from 5 to forty. However, in many instances the methodological consideration appearing in these other criteria are a minor part of the more general criteria (e.g., specification of the management plan). As a consequence, the liberal definition is likely to be an upper limit on the relative importance of methodological considerations.

The number of points that are listed as the minimum necessary to be considered for funding range from 30 to 70. Given these conditions, it is possible that projects which are methodologically unsound could easily receive funding or at least consideration for funding. Previous experience in evaluation research has pointed out the importance of initial planning in order to conduct a successful study. The weighting scheme, as it is specified in the regulations, seems to impede early consideration of methodological issues.

Several options seem feasible. Given that these types of discretionary grants are made available for testing new, innovative ideas, it seems reasonable to increase the emphasis on the evaluation plan to ensure that good ideas are adequately examined. This, of course, implies that there is sufficiently trained personnel available at the proposal preparation phase of the application process (at the LEA and SEA level). In the event that there is not sufficiently trained personnel, an evaluation negotiation phase could be imposed where the grantee could obtain federally sponsored technical assistance. Once the evaluation plan is negotiated, continual monitoring of the evaluation process should be carried out for the duration of the contract. This is especially necessary when an LEA has to rely on outside contractors and/or when local evaluation capabilities are less than sufficient. As the application process is currently structured, the SEA is supposed to receive a copy of the proposal prior to the time the proposal is submitted. The regulations do not, however, specify who is to conduct the review and it seems reasonable to state explicitly that an evaluation specialist and program specialist should provide comments on the methodological and educational quality of the proposal. Our interviews with state level personnel suggest that the review is usually carried out by substantive experts and routine examination of proposals by state level evaluation personnel is the exception rather than the rule.

5.3 CRITICAL EVALUATION ISSUES[2]

During the short history of evaluation research, critical issues that threaten the quality and integrity of evaluation efforts have become apparent. This section considers four topics bearing on the evaluation of new programs and components of programs.

The first section considers implementation of programs, the time it takes to put prototypes into the field. The second and third address design issues especially the use of randomized field experiments to plan and evaluate new programs. The last section concerns criticism and reanalysis of the results of evaluation, primarily at the national level.

Program Implementation

Difficulties in Implementation are Chronic. It is not difficult to identify instances of notable discrepancy between program plan and program activity. Early Title I funds were used to supplant, rather than to supplement, existing funds as the law requires. At least some summer compensatory educational programs did not receive funds in time to spend them correctly during the late 1960's. More recently, the Study of the Emergency School Assistance Act suggests that analogous discrepancy between statutory requirements and local assignment of funds characterized a few school districts' operating programs under the ESAA. That it is not only fiscal aspects of a program which may be problematic but also (more often perhaps) staff level and staff activity, is clear from case studies of Performance Contracting, Follow Through, Planned Variations, Middle Start, and others.

That the problem emerges in smaller scale efforts as well as in the large is evident from recent work at Ohio State on High-school Internship Programs, evaluation of media-based instruction such as Sesame Street, studies of variability in home-based instructional programs, and others. The persistence of such problems suggests that in tests of new programs both control and program groups be routinely monitored for dilution of programs since estimates of program effects under these conditions will not be accurate. For massive complex programs, such as Title I, high quality sample surveys such as the Sustaining Effects Study have been informative. But we have uncovered no formal federal policy which would provide for periodic sample information of the same or higher quality.

The Difficulties of Implementation are not Confined to Education. Problems in assuring adherence to a program plan and in assaying the nature of adherence are not confined to educational innovation, of course. In early experiments on reducing retrolental fibroplasia, nurses were often unwilling to cooperate with researchers in depriving premature infants of a highly enriched oxygen environment, then considered beneficial. Studies later demonstrated that infant blindness is caused by the routine high-oxygen treatment. Analogous difficulties have been encountered in other areas of social experimentation -- examples from engineering, biochemical research, and pharmaceutical control are not difficult to find.

Part of the Problem Appears to be Time for Implementation. Estimates of the time required for implementation and stabilization of new programs are crucial to the agency executive and program managers in understanding limits on progress as well as the usefulness of outcome evaluations. But available information exists more often in the form of expert judgment than elaborate, verifiable evidence. Garry McDaniels of the Bureau of Education for the Handicapped, for instance, regards one-year implementation periods for new programs as absurdly inadequate and maintains that some sponsors of Follow Through models required two to three years as an absolute minimum for implementation of new programs lifted directly from the laboratory. According to NIE's Lois-ellin Datta, the time allocated for introduction and stable operation of Planned Variations -- two years -- was barely adequate. Alice Rivlin and Michael Timpane observed that despite the fact that High/Scope sponsors were the most sophisticated of the Planned Variation sponsors, five years of development in the actual sites were required before they could say the program was fully developed. The System Development Corporation's John Coulson puts the time at three years for stabilized ESAA programs. Emrick's case studies of school district adoption of innovations suggest that implementation generally takes more time than expected. "Virtually all" were in their second year of involvement with the project and the first year of implementation (p. 119).

We have been able to uncover no more precise assessments than these in the published literature, and believe that the absence of coherent information on the topic is a serious problem.

Attention to Implementation at the Local Level. According to the UCLA study of large school districts, only about 20% of the directors of evaluation rank implementation assessment as one of three most time consuming efforts. This stands in contrast, for example, to the 70% who regard assessing results of programs as very time consuming, or the 66% who spend a great deal of time on measuring student objectives. Their review of evaluation reports issued by school district evaluation units suggests that less than 17% of such reports cover program implementation -- although most reports do concern a particular program.

Defining and Measuring Implementation at the Local Level. In evaluating the National Diffusion Network, John Emrick and his colleagues at American Institutes for Research identified some severe problems in understanding what "adoption of an innovation" means. The issue is critical since NDN's goal is to facilitate adoption of tested innovations by LEAs, and it should be evaluated partly on the basis of achieving the goal of adoption . The adoption is related to the issue of implementation. His suggested indicators of adoption are (a) scale of adoption, how much, at what level, or how many components of the original innovation are implemented, (b) novelty of adoption, to what extent is the innovation much different from what went on before, (c) fidelity, or degree to which the implemented version of an innovation matches the original innovation.

As a practical matter, AIR defines scale of adoption in several ways, including the proportion of people who admit implementing only part of the complete package rather than the whole innovation. In assessing fidelity, AIR suggests using modification rate -- the proportion of people who say the innovation must be changed in reasonable ways to be compatible with the system.

Definition and Measurement at the National Level. Establishing whether a program is implemented, whether it has stabilized, and what its level of implementation is relative to a standard is a legitimate matter for evaluation policy. It is not essential that the pertinent questions be answered precisely at the oversight level, though there may be considerable need for thorough measurement in the project-level evaluation. In point of fact, the GAO makes assessment of implementation explicit in its own policy on oversight, and this continues the agency's durable tradition of verifying that programs exist or fail to exist. The need to treat implementation as a policy issue has been reiterated by academic advisors to government, such as Peter Rossi and Howard Freeman, and agency executives such as Michael Timpane.

Microassessment may also direct attention to the stage of a program's development. It is clear from Follow Through, for instance, that direct transferral of laboratory programs to the field is likely to engender major implemtation problems. It could also include the attention to expenditures, manpower assignment, samples of transactions and delivery of tangible goods, which typifies the administrative audit. University of Washington's Richard Elmore summarizes the view, arguing that it is sensible to measure "dimensions of classroom activity before a program's introduction as well as after... to identify major characters in the process and their roles, and enumeration in detail of what is to be implemented."

In large-scale evaluative surveys of existing programs, the analyst's interest lies in measurement of implementation rather than actual control. Experts such as Spady, conclude from school resource research that "crude and deceptive measures of supposedly relevant resource variables" are a distinctive problem, and that "the school resources tapped (measured) in a majority of studies are crudely or even unreliably measured, stress quantity or mere presence over quality or mode and degree of utilization ... and are really only proxies for resources which actually reach children." The Pettigrew/Green and Coleman debate about the large-scale data used in analyses of white flight following school desegregation suggests that the indicators of desegregation are not always well understood, despite their ostensible pertinence to assessing impact, nor are they uniformly reliable.

Microassessment of Program Implementation. Despite the educational researcher's conscientious attention to measuring children's responses to programs, such as achievement, and despite the manager's equally conscientious attention to programs' operational problems, the notion of developing and using methods of measurement to gauge implementation is relatively new. Without such methods, it is unlikely that future innovators will be able to avoid the less visible sins of their precedessors, to define "replication" any better than we do now, or to construct more credible estimates of program effect. At best, that quantitative indicators of program implementation can be exploited to design more sensitive evaluations and that the effort needed to develop those indicators must be catholic, involving managers, theorists, and measurement specialists, are also clear.

That it is possible, with sufficient financial support, to be more thorough in characterizing programs is evident from research on teacher-child interactions in classrooms at Arizona, Stanford, Texas, and elsewhere. Evaluations of viewing time and other factors affecting receipt of educational television programs appear in Sesame Street, Electric Company Feeling Good, and Freestyle. It is not unreasonable to expect to do a better job in the territories traditionally claimed by economists. John Coulson's descriptions of the ESAA evaluations suggest that Systems Development Corporation has made notable advances identifying the people who are informed about budgets within school districts, in eliciting clear information, and in identifying subtle stereotypical flaws in the process. Critics such as Weinberg assert that even the Coulson work could be improved considerably by focusing on expenditures rather than on budgets.

The need for better measurement, if not desperate, is remarkable. Pleas for attention to the problem have been made by policy analysts in reviews of planned variations, such as Elmore at Washington, by federal agency managers such as Datta at NIE, program developers such as Weikart and Banet, at High/Scope and some advisory groups, such as the Social Science Research Council.

An implication of all this is that in oversight policy, programs ought to be identified by more than pious promise, rhetorical label, or contract statement. More specifically, better information on temporal character of implementation, stereotypical misdirection, and other crude dimensions of implementation ought to be collected systematically. What is less clear is how the job can be done effectively. Options under such a policy include, for example, sample surveys as in the Sustaining Effects Study, special investigations such as the GAO's examination of Follow Through, and uniform reporting which avoids or at least recognizes misreporting. Emrick and others recommend that "criteria for assessing, documenting and reporting the status and quality of adoptions be developed" for each new program in discussing the National Diffusion Network, but the advice is sensible for at least some new programs.

The support of research on methods of assaying implementation is essential if we are to understand (1) how to construct less expensive methods of observation (the existing ones are often expensive); (2) what impact the observational processes have on implementation (e.g., they may foster implementation in some cases or the illusion of implementation beyond the readily observable dimensions, including quality of student-teacher interactions rather than just frequency).

Evaluation Design

By evaluation design here, we mean a plan for assigning individuals to programs or to program variations, collecting data, and analyzing results so as to produce reasonably accurate estimate of the relative effects of programs. This stress does not mean that other types of evaluation are any less important in a particular case. Questions of how many students are served and what services they receive appear to be no less important in the Congress's view.

Quality of Design. The quality of an evaluation design for estimating a program's effects will influence the quality of evidence about those effects. To the extent that state-of-the-art design is ignored, evaluative data on program effects will be ambiguous at best and misleading at worst.

The idea that evaluations ought to be designed and begun before a new project's installation was novel to most funding agencies 15 years ago, and it is still a novelty for most private foundations. Rather, the emphasis was on after-the-fact analysis--post mortem or opinion survey--with its attendant shortcomings. It is remarkable that attention to evaluation design has been focused in the intervening years to the point where policy emphasis on design is evident at the Assistant Secretary level in DHEW judging from Henry Aaron's testimony before the Committee on Human Resources. This does not mean that most designs employed in recent studies have been perfect. It does mean that the poor quality of earlier evaluations is better documented, the design-related reasons for poor quality better understood, and there is some political pressure to eliminate or control inept designs.

The justification for attending to evaluation design stems partly from reviews of educational program evaluations generally and commentaries about inept evaluations at the school district level (e.g., Hawkridge, Chalupsky, and Roberts on Title I in 1968). Others reiterate the lesson based on case studies of selected educational program evaluations and medical projects (e.g., by Gilbert, Light, & Mosteller at Harvard). In extended reviews of single projects such as Head Start by Rivlin and Timpane, for example, they observe that "researchers found out, late and the hard way, about the costs that design flaws and compromises. . .exact in the form of weak or uninterpretable results" (p. 11). More rudimentary design problems helped to make data from Performance Contracting evaluations nearly useless for estimating program effects, judging from Gramlich and Koshel's analysis at Brookings--though this evaluation was notably informative in other respects. Broader critiques, such as Bernstein and Freeman's make the same point. Such studies are themselves not flawless (see Abt & Abt, and Wortman, for example), nor do they accord much publicity to well-designed evaluations which, though not in the majority, are no less important. They are clearly a useful vehicle for understanding that some evaluations have been inadequate and for understanding why they were inadequate.

The problem of assuring good evaluation design is not confined to the education sector. The GAO's early criticism of evaluation of the Experimental Housing Allowance Program was based partly on GAO's judgment that the design was flawed. In medical research, Chalmers, Block, and Lee provide clever evidence, comparing results of randomized experiments against those of nonrandomized tests, to make a similar point. Silverman's description of 20 years' work on blindness among premature infants, which appeared in Scientific American, is an especially nice model from medical research. It identifies the ways in which poorly designed tests were both helpful and dangerously misleading in early work, and the later resolution, using randomized clinical trials, of controversy over the causes of blindness. Partly because education, unlike medicine, is a rather recent interest for the science writer or historian, long-term case studies of the consequences of poor design decisions in educational, research are more difficult to find.

One immediate implication is that attention to quality of an evaluation design should be incorporated routinely into policy on program development and evaluation. This is not to imply all evaluations must be of high quality. It is to say that choices of tolerating minimal evaluation or to support a high-quality one ought to be recognized as such.

The idea that new programs should not be designed and put into the field apart from their evaluation is explicit in reports of the GAO on specific projects such as Follow Through and in more general guidelines on oversight policy. The effort to infuse attention to quality into Congressional review and into evaluation law is also reflected in legislative staff activity; for example, early guidelines for Congress developed by Foskett and Fox, Fox's evaluation seminars, and Franklin Zweig's more recent efforts to assist the U.S. Senate's Committee on Human Resources. Within agencies, it is difficult to distinguish between individual professional preference and general agency policy. For example, NIE's Lois-ellin Datta stresses the roles played by former NIE directors Thomas Glennan and Harold Hodgkinson in supporting rigorous assessments of NIE's Career Intern Programs. At least one divisional manager at NIE has been consistent in encouraging high-quality tests for the work experience programs despite early resistance among some contractors to the policy. At the level of the Office of Program Planning and Evaluation in the Office of Education, some directors have been strenuously arguing for quality in design, basing their advocacy partly on the controversy over outcomes of the Head Start evaluations (Evans, 1974; U.S. Senate Subcommittee on Oversight Procedures, 1976). At the Assistant Secretary level, Aaron (1978) recognized more generally the lack of uniformity in quality of evaluations supported by DHEW, suggesting that they run from "bad to excellent" and that the outcome depends heavily on thought invested at the design stage.

Policy which emphasizes quality in evaluation design cannot be expected to implement itself or be adopted immediately. It is for this reason that improvement in caliber of staff responsible for review of design is crucial. Efforts to recruit that staff have, according to Aaron, been vigorous at the level of Assistant Secretary for Program Evaluation in DHEW. Judging from the conversations with a few members of that staff, the effort has had remarkable results. However, we have been able to uncover no serious academic attention to staff improvement in this sector or in other federal agencies, such as the GAO, where staff improvement has had high priority.

Randomized Field Experiments

Randomized experiments, in which children or classrooms or schools, are randomly assigned to one of two or more alternative programs which have the same aim, are a promising vehicle for obtaining a fair estimate of the relative effectiveness of the programs. The usefulness of randomized tests in principle is generally not at issue. When experiments are conducted properly, orthodox theory of experimentation guarantees that long-run estimates of effects will be unbiased. The cost of this increased assurance of interpretable results is a greater demand on managers of programs. The demand includes executing and temporarily maintaining the randomized assignment and other necessary cooperation with the evaluator.

Argument about the use of the design more frequently concerns the idea that randomized experiments are rarely feasible in field settings. "Rareness" and "feasibility," however, are infrequently specified by the government policy group such as the NIE's Task Force on Resources Planning in 1974, or by the individual analysts such as Horst, Tallmadge, & Wood. It is true that, although the design is not new, its application in evaluating educational and other social programs is relatively novel. But novelty does not establish lack of feasibility, and a notable, if not large, number of field experiments have been mounted.

The most recent examples include evaluations of: parts of the Emergency School Aid Act (by Coulson), a subset of career education programs (NIE; Datta), Middle Start program run at Oberlin College by Yinger, Ikeda, Laycock, & Cutler, educational TV programs in health developed by Children's Television Workshop (Mielke & Swinehart; Minor & Bradburn) preschool education (Bogatz & Ball), primary education (Ball & Bogatz), radio-based mathematics instruction (Searle, Friend, & Suppes), and even grade retention (Jackson). Oliver Moles at NIE has managed to implement randomized tests of programs that were designed to reduce distruptive school behavior. Welch and Walberg at Illinois are among the embarrassingly few researchers in any discipline to have mounted randomized experiments to test a dissemination/utilization strategy. (George Fairweather and Louis Tornatzky have done remarkable work in mental health). Rickel, Smith, and Sharp of Wayne State University have executed remarkable rigorous tests to establish the effectiveness of a preventive health care for preschoolers with behavioral and emotional problems served with Title I program funds in the Detroit School District. The Cali, Colombia, tests on education and health programs for malnourished and educationally deprived children are a milestone in the developing countries (McKay, Sinisterra, McKay, Gomez, & Lloreda). Partly to capitalize on problems and hard lessons in the original planned variations study, USOE issued an award-winning RFP in 1976 for higher quality evaluation of few planned variations programs for youth. The project was terminated by USOE for reasons lying in the program developing arena as well as in the design arena. Most such field tests have been moderate in size, Emergency School Aid Act (ESAA) being an exception. Others have not been successfully executed even if they were well designed by technical standards, as one should expect. Boruch, McSweeny, and Soderstrom's bibliography of attempts to mount randomized tests includes both successfully executed ones as well as disasters.

The state of the art in executing field experiments is developing rapidly. One of the clear lessons of the past 10 years is that solutions to managerial, institutional, and other problems are at least as crucial as the statistical and analytic ones. The incentives for experimentation have also been better articulated, partly on the basis of case studies of nonrandomized evaluations. This includes the arguments in both academic and policy quarters that the original Head Start evaluations inadvertently produced biased estimates of program effects (Campbell & Erlebacher; Evans). And it includes observations by Smith and Datta that the results of some quasi-experiments, such as Head Start Planned Variations, are unlikely to provide clear evidence about the models' relative effects even if they are informative on managerial grounds. The special problems engendered by the need for a strict regimen in assigning individuals to programs and the conditions under which randomized trials are feasible, are better understood as a result of case studies on evaluations in education and other areas (Riecken and others). The legal status of random assignment rules has, until recently, not been clear. Three judicial decisions on the legitimacy of randomization in testing programs are reviewed by Breger; those decisions and the underlying theory are generally favorable for the judicious experimenter.

The implication of all this for policy is that randomized field tests should be regarded as a legitimate option, if not always the preferred one, for testing new social programs when estimates of relative program effect are important enough to justify their cost. Consistently well-executed field experiments in education are likely to require greater changes in legislative posture than are those for medical research. As long as political feasibility, rather than interest in impact, is the primary rationale determining evaluation design, demonstration projects, rather than experimental tests, are likely to remain the norm.

Quality of Analysis

One of the main results of surveys and interviews with Congressional staff members is that they are concerned about credibility of the findings of evaluation at the national level, and to a lesser extent at the local level. The matter is a red herring in one respect and a legitimate issue in others.

The specious aspect concerns assaults on evidence when it turns out not to be favorable to one's position. This is especially true of evaluations which are dedicated to estimating program effects. It has clearly occurred in recent controversy over evaluation of bilingual education programs; it occurred as well in early Title I evaluations. Simply put, not all criticism is competent. The legitimate feature of the problem is that evidence is often incomplete and its quality uncertain. To the extent that the state of the art in analytic methods is unsettled, then some argument about the proper analysis and proper inferences is warranted. For feasible evaluation designs there is considerable room for debate over proper analysis and conclusions. This is as chronic a problem in medicine as it is in education.

Critique and Secondary Analysis

One of the major implications of contemporary experience is that major program evaluations should be subjected to formal critique and, where necessary, reanalysis of original data. By critique here we mean balanced and independent appraisal of the evaluation not just negative criticism. In particular, the results should be reexamined by individuals independent of the original investigators. "Findings" here include data, analysis, conclusions, and recommendations. The implication holds for any type of evaluation used as a basis for policy.

The reasons for this policy and the forms it may take have become clearer over the past five years, partly because of actual reanalyses of data stemming from evaluations of Sesame Street, of Head Start, the Equality of Educational Opportunity Surveys, Follow-Through, and others. The most recent illustration stems from a multi-year Rand investigation of federal programs supporting educational change. One part of the report criticized technical assistance rendered by consultants to a school district on program development. This aspect of the report was widely publicized and widely believed in and out of government. We are aware of only one critique of the study, by Lois-ellin Datta of NIE. It shows nicely how the data presented in the report do not support such a simple conclusion.

Regardless of whether Rand's conclusions are justified, the point is that no serious formal criticism of the report was undertaken before or during its release. There are several important reasons for considering formal critique or reanalysis, regardless of the skill and integrity of the original investigators. If one espouses the view that most effective internal evaluators should be benign skeptics, the need for external analysts who are benign in different ways, if not less benign, is justified on scientific and political grounds. To the extent that independent secondary analysis is routine and visible, it may impede the corrosion of credibility of large-scale program evaluations of the sort described by McLaughlin in early Title I programs. The more durable reasons for secondary analysis are scientific: Verifying quality of information and identifying both egregious and sophisticated errors in analysis are basic to that interest. That egregious errors are sometimes not hard to find is evident from GAO's reexamination of early Follow-Through evaluations for instance. Secondary analysis is an economy-minded notion in the sense that an expensive data set is made to work repeatedly at low cost, if we judge by the Head Start experience and Lois-ellin Datta's examination of the reanalyses, regardless of its quality.

This is not to imply that all reanalyses of raw statistical data will be worthwhile, since the worst of cases can be identified easily in critique by external reviewers. Nor will all reanalyses produce new findings: Indeed, merely certifying that original estimates of program effect are reasonable may be sufficient. Finally, reanalysis may be no more illuminating than the original analysis. This is still useful in the simplest sense: Two ambiguous interpretations may help to verify that things are indeed as bad as they seem to be.

The view that reanalysis of raw data should be regarded as a legitimate option in overseeing evaluations has been adopted in GAO's draft monograph on appraising impact evaluations. Within some divisions of agencies such as NIE, for instance, de facto policy of reanalysis exists. We judge this from conscientious maintenance of data generated by evaluations of the Emergency School Assistance Act, career education programs, and others. Some exploration of the topic as a matter of internal policy has been undertaken at the federal level by Virginia Koehler, occasionally at the state level judging from Powell at Colorado, and by academic analysts. NIE appears to be the only agency until recently to have made a substantial fiscal investment in supporting research on reanalysis of program evaluation data. But the general policy of supporting secondary analysis has not been officially sanctioned nor explicit at NIE, USOE, or other agencies, such as the Agency for International Development, which evaluate new educational programs.

Data stemming from surveys and evaluation activities by Parent-Teacher Associations, community groups, and the like are rarely subjected to any real reanalysis despite their import for local policy and occasional import for national policy. But the justifications given for reanalyses of federally supported evaluations appear no less pertinent to this arena. It is not unreasonable to expect that federal policy can serve as a model in this area for state agencies and legitimate interest groups at the local level. To the extent that policy on evaluation and on support of these groups' applied research also stresses secondary analysis, the quality of the product is likely to improve.

Storage of and Access to Information for Secondary Analysis

The Freedom of Information Act guarantees access to a variety of information useful for evaluation at the federal level, and some state laws provide similar assurance at that level. As principle, this can help to justify access. Despite the law, it has been difficult at times to acquire data used for program evaluations. The difficulty is certainly not confined to education, judging from the illustrations in crime deterrence, biochemistry, and medicine. The difficulty is influenced by flawed program management, poor record keeping practices, professional jealousies, and the bald fear that an analysis will be found incompetent. Formal policy cannot be sensitive to each problem, but it can be used to reiterate the message carried by law and to routinize the opportunity for conscientious reanalysis. The only written statement of internal policy we have been able to locate has been developed by the Law Enforcement Assistance Administration (LEAA). The policy hinges on the requirement that statistical data generated in LEAA-supported research be turned over to a special LEAA-supported archive at the University of Michigan when research reports are submitted. It is consistent in spirit with preliminary papers developed at NIE and with general policy developed by the President's Federal Statistical System Project.

Adoption of the idea that data ought to be made available for reanalysis creates choices about whether to support decentralized archives, whether to screen data sets for their political or scientific value, and so on. The National Archives have made explicit their interest in storing all machine-readable evaluative data, and this will help to attenuate management problems in the policy's implementation. Vehicles other than the National Archives are worth considering since no quality control system for screening of data or documentation will be undertaken by the Archives.

Privacy and Secondary Analysis. When secondary analysis depends solely on statistical data from which individual identifiers have been removed, there are usually no special privacy problems engendered by the data's disclosure, at least for individuals. A class of problem cases, in which deductive disclosure of information on individuals is possible, is so small and so idiosyncratic for major studies that, as a matter of policy, it can be accommodated by ad hoc review of the special cases.

The more likely problem, judging from recent arguments over access to data in Lora vs Board of Education of the City of New York is that (a) claims of the risk of deductive disclosure will be made, (b) no clear standard and no formal explication of the evidence for the claim is published and (c) the claim will be entrained in other, more important reasons for refusal to disclose (e.g., protecting institutions, such as schools, against charges that they have violated individual civil rights).

For secondary analysis of identifiable records, the privacy problem is rather more crucial. It has been subjected to investigation by oversight agencies such as the GAO and advisory groups such as the Social Science Research Council, and legislatively mandated review bodies such as the Privacy Protection Study Commission. The first of two major conclusions stemming from this work is that there is a large variety of procedural methods for linking data from different record systems and for checking the quality of records, without abridging privacy rules governing each source of information. Those methods have been tested in education, epidemiology, and other applied research. They should be considered by independent researchers and to oversight agencies such as the GAO where rules limit access to identifiable records. Second, where identifiable records are necessary for legitimate research, disclosure of those records to the researcher should be permitted provided that stringent conditions be met. Those conditions include conscientious review of the process, clear agreement of the research function and nature of disclosure, and prohibitions and sanctions against violations of agreement and redisclosure.

5.4 CONSTRAINTS ON THE QUALITY OF EVALUATION

Technical factors are not the only influence on the quality of evaluations. Three other factors are considered here: The appropriateness of limiting communcation between agency and Congressional staff; Federal Clearance Requirements; and Independence, integrity and candor in reporting.

Direct Contacts Between Congressional Staff and Agency Staff

Over the last five years, memoranda have been issued by the Office of the Secretary of DHEW have periodically concerning contacts between Congress or the Congressional staff and federal agency staff. The theme has generally been that such contacts in the absence of an agency legislative liaison staff are not desirable. There have been official demands that liaison accompany any staff.

The justification for restrictions on contact centers around the probable difficulties engendered by numerous agency staffers dealing with the Congress directly. Positions of the Department may be misrepresented or misinterpreted. The difficulty is not confined to Education of course. We recognize that there is a need to avoid program initiated "lobbying" but take issue with the uniform application of this policy as it is applied to the process of conducting evaluations. Producing a timely, high quality and potentially useful study, requires unimpeded access to relevant Congressional staff to ensure that issues are properly addressed before and during the execution of the study. While restricted contact may be appropriate for some Departmental issues, it is counterproductive when evaluations are concerned.

The major side effect of this institutional policy appears to have been impeding the agency's ability to understand diverse Congressional interests in evaluation, to impede development of reasonable basis for Congressional staff views of agency evaluation activities, to impede the negotiations which are part and parcel of any major evaluation enterprise. Within OED, these sentiments have been expressed by two of the three divisional directors to whom we spoke. And the public remarks by knowledgeable Congressional staff reflect the same spirit. John Jennings, for instance, has pointed out flaws in OED activity; misunderstanding the law and its requirements, creating an evaluation plan when the evaluation itself has no chance of being used by the Congress, and others. Each of these could be avoided with reasonably unrestricted conversation.

The result of formal restrictions is that agency staff have not felt free to call Congressional staff to ask questions, to verify their own interpretations, to clarify problems in what to evaluate and how to evaluate it. This is compounded by inability or unwillingness of some Congressional staff to initiate a conversation with the agency staff.

We believe it is a sufficiently complicated and serious problem to warrant attention from Congress and the Department of Education. Major restrictions on conversation about evaluations which the Congress demands and the Department must provide are absurd when the utility of the evaluation depends so heavily on mutual understanding.

Federal Clearance Requirements: Reports

Proscription against disclosure of information collected during an evaluation without formal clearance can be and has been an impediment to evaluation. It can for instance, truncate the government's access to manpower for evaluation. In the worst case, it may lead to delays in providing information.

To illustrate, consider directives issued by the Secretary of Health, Education, and Welfare in 1978. Until then, evaluation reports had been issued routinely by the Office of Evaluation and Dissemination to authorization and appropriations committee members, and relevant members of Congress and the agency. Early in 1978, Secretary Califano evidently became "aware that procedures for transmitting reports needed to be more clearly defined," and made his recommendations explicit in an April 10, 1978 memo. The immediate cause for concern appears to have been queries to the Secretary on information released earlier. The action taken in a memo from Califano to DHEW office heads was to require that all reports and evaluations be reviewed by the Executive Secretariate prior to release. The single exception noted involved reports specifically mandated by Congress to go to Congress directly. The requirement was put into effect for final reports.

The memo also appears to have been instrumental in strengthening a requirement that contractors abide by the same rules. In particular, Article 28 became part of the boiler-plate for evaluations executed with USOE funds. The article requires that the contractor not disseminate data without written consent of the contracting officer.

The general requirement for clearance by the Executive Secretariate led to notable delays in release of reports produced by the Office of Evaluation and Dissemination. Judging from the 13 reports submitted for review between January 1979 and May 1979, the delays range from 18 days to 133 days; a quarter of the reports were delayed by over two months. Moreover, only one report of this collection was returned to the originating division for revision. According to a memorandum from Jim Pickman of OED to Rick Cotton of the Secretary's office, the remainder received no notable modification as a result of review by the Secretariate.

At least one major effort was undertaken by OED executive staff to assure that review by the Office of the Secretary had an opportunity to review reports within a specific time. It did not succeed. The current Office of the Deputy Assistant Secretary of Evaluation and Program Management, has however, instituted a new clearance system and established a clear limit of 10 days on release of reports. In particular, the offices of the Secretary, Under Secretary, Assistant Secretaries, and Deputy Assistant Secretaries have "ten working days to review the summary and, if they wish, the contractor's report," in accord with a directive from the Deputy Assistant Secretary for Evaluation and Program Management.

Article 28. This clause prevents a contractor from discussing data produced in the execution of the study without permission of the contracting officer. Interpreted by at least some lawyers, this means that information generated during an evaluation cannot be presented, for example, at professional meetings prior to agency review. For the uninformed, it implies that results cannot be discussed without prior permission. We have been told by executives of two major contracting organizations that their staff members merely send a copy of the paper they plan to present at a professional meeting to their project monitor (not to the business office of course) and routinely receive permission to present results. The basis for this agreement appears to be the reasonable expectation that professional presentation will help illuminate strengths and weaknesses of the information. Often the routine approval is provided because of trust in the capability and integrity of a contractor selected in a competitive process. Apparently, the article affects some academic institutions in the same way.

However, some well regarded academic institutions do not treat the restriction on information as informally. As illustrated in the case study included here, Northwestern University refused to agree to the pertinent article in contract negotiations for research leading to this report. The National Academy of Science's legal counsel raised a similar issue in negotiations for a parallel Committee on Program Evaluation. The argument to many is an unnecessary one, given that no funds for evaluation in education have ever been withdrawn for violating the article. But it is sufficiently real in principle to warrant argument.

Independence and Integrity

Federal Agencies. The question of whether the agency responsible for reviewing a program should also be responsible for evaluation is an old one. It is implicit in the early history of the U.S. General Accounting Office, and more recent history of the Office of Evaluation and Dissemination at USOE. The arguments for and against the approach can be put tersely.

The advocates of conducting evaluation independent of program staff argue that evaluation reports are more likely to be candid and fair if they are conducted by individuals outside the jurisdiction of the program. On the other hand, advocates of evaluations being carried out by program staff, rather than by independent evaluators, maintain that program staff are more knowledgeable, better informed, and that evaluations can be more useful when done by program staff.

Case Studies of Impediments to Timely Research: Article 28

Public Law 95-561 required that an evaluation of educational program evaluations be undertaken during 1979-80 and that a report be submitted to Congress by the Commissioner in July 1980. The law further required that the Commissioner of the U.S. Office of Education make available funds to an independent agency to undertake the study. Independence was emphasized in the USOE's work statement to assure fair examination.

John Evans of the USOE's Office of Evaluation and Dissemination (OED) invited proposals for the research from several independent researchers. On July 5, Boruch discussed the matter with Evans and John Jonas of Representative Holtzman's office. A formal proposal was submitted on August 31. The discussions with Evans and Jonas repeatedly stressed independence of the investigator.

Northwestern University was provided with a copy of the proposed contract on September 25 by the USOE's contracts Office. Officials at Northwestern read the business section of the proposal, finding two offensive clauses. The first, involving time commitments of the principle investigator, was worked out. The second clause involved a more fundamental conflict. The clause said:

> ARTICLE 28. DISSEMINATION OF DATA. No subject data as defined in the "Rights in Data" clause in the General Provisions may be disseminated with the express written consent of Contracting Officer.

University counsel objected to the clause on grounds that it violated academic freedom of inquiry and expression. The principle investigator concurred in a telephone conversation with Charles Seibert, Assistant Director of Northwestern's Office of Research and Sponsored Projects. The University signed part of the contract but returned the offending portion, unsigned, saying that the clause would have to be removed before the University could commit itself to the work. The partly signed contract was sent to the USOE Contracts Office on September 27. During September 28, Seibert spoke to staff members of the Contracts Office, and was told that the Office's position was immutable. The clause would stay. Seibert then assumed the contract could not be consummated then, and since the fiscal year ended on September 30, a Sunday, it could not be consummated at all.

On Monday, October 1, at 1:50 p.m., Boruch called (OED) to verify the problem and to determine if indeed the Holtzman Project had gone up in smoke. The OED announced that the Contracts Office had deleted the clause and signed its portion of the contract over the weekend after some argument. We were both told that the National Academy of Sciences had taken the same position as Northwestern's independently in negotiations on another contract, and the clause had been deleted for NAS also. Boruch called the Contracts Office to confirm this, and did so.

Exactly the same arguments appear in international debates by directors of statistical agencies. Recent meetings of the International Statistical Institute in Warsaw, for instance, focused heavily on whether those agencies should be responsible for analysis of data as well as data collection, or whether routine analysis by independent agencies is warranted.

In this Project, the question emerged in interviews with federal agency staff, and was implied often by Congressional staff. The debate is complicated because bureaucratic territory and political power are at issue. An agency may prefer internal evaluation partly because budget can then be increased to conduct the evaluation.

Federal Contractors. We have encountered no evidence of falsification of data and no complaints about the matter for large federal contractors. Arguments do arise, however, over technical analyses and what can be inferred from the data, and occasionally over the extent to which a contractor's recommendations are based on the evidence. These seem to us to be more a matter of competence and reasonable differences in values than deliberate deception.

The questions answered by the federal contractor are normally posed by the federal agency staff. If those questions are inappropriate or narrow, the answers, regardless of their accuracy, will be partial at best. We have found contractors who believe agency staff asked the wrong questions. For example, Clark Abt maintains that the focus of research on Follow Through should have been on identifying remarkable successes. But we've encountered no real questions about integrity of OED staff in performance of their duties.

A number of mechanisms are already in place to enhance quality of work. To some extent these make integrity a moot issue. The peer review system does operate for any grant or contract issued by OED and NIE, except for regional laboratories and centers, and Congressionally mandated projects. Review and advisory boards do exercise oversight on quality design, execution, and anaysis, though this varies from project to project. In principle, data are available for reanalysis and indeed a variety of reanalyses have been undertaken.

The Local Level

Independence at the Local Level. In large school districts that have evaluation units, the units depend most heavily on school operating funds for their support. They do however depend on federal funds directly. To illustrate, the Austin School Districts Office of Research and Evaluation was created in 1973 on the basis of an ESEA Title III grant. District funds were gradually used to build up the unit and current state funding permits hiring research personnel. Fiscally, then evaluators are not entirely dependent on their school district resources, neither are they completely independent.

Administratively, evaluation units are responsible directly to superintendents rather than to program directors. In this sense they are independent. However, they provide service to programs and their likelihood depends at least partly on cooperation of staff and their relations with institution staff.

Administrative independence was not identified as a problem in the districts in our site visit sample: Most of the evaluation units reported to the central administration either directly or through an intermediary unrelated to the programs. Results from the CSE study, suggest that administrative independence has either been achieved or is not a serious problem for most large districts. For districts without an evaluation unit, administrative independence is less frequent. The options here for promoting independence are to encourage the establishment of evaluation units for moderate size LEAs through the use of federal funds, as in the case of the Austin School District. For smaller districts, there simply are not enough pupils eligible for federal assistance to warrant establishing a unit, the use of contractors might be urged or numerous small districts might pool their resources. Consortia have also been offered as an option to ensure the availability of sufficient monies to attract high caliber outside contractors.

Lack of Fiscal Independence for Evaluation. One problem that we encountered in our site visits and phone conversations concerns the issue of fiscal independence. While evaluation units in most SEAs and some LEAs are administratively independent of programs, they usually do not have control over the allocation of evaluation monies. This responsibility typically lies in the hands of program directors or administrators of federal programs and may invite conflict. For example, in Site G, the Director of Federal Programs controlled the Title I evaluation monies. Consequently, his approval of monies for Title I evaluation efforts and even evaluation staff attendance at Title I workshops and professional conferences was required. Since he was not particularly receptive to evaluation, this consent was not always forthcoming and resulted in restriction of opportunities for staff training, professional activity, and improvement of evaluation activities. The conflict ultimately contributed to the resignation of a highly trained Title I evaluator. Another district, although permitted by the SEA to allocate up to 5% of Title I monies for evaluation, was embroiled in disputes between the Directors of Title I and the Evaluation unit. While the Director of Evaluation wanted to improve the quality of Title I evaluation efforts by engaging in activities which went beyond compliance, the program director was unwilling to allocate the maximum set aside permissible. Consequently, these evaluation units which wanted to improve evaluation efforts or respond to other district requests were hampered in their efforts to do so.

This is a complex issue and the need for alternative allocation mechanisms is exemplified by the situation in Site B. Here fiscal dependence did not limit the nature of evaluation activities but rather allowed greater flexibility. Prospective budgeting is often difficult in evaluation as unanticipated problems arise in data collection and other such tasks. In this district, the Director of Federal Programs did award additional monies for evaluation when problems arose and a strong case was presented. This would not have been possible when fiscal independence existed. However, this partly was conditional upon the Director of Federal Programs' positive attitude toward the evaluation process. Relying on the presence of positive attitudes to ensure cooperative relationships is insufficient--other options need to be entertained.

One possible way of ensuring fiscal independence is the strategy implemented in State I. Here the legislature reappropriates the federal money received and allocates it to the programs and evaluation unit separately. Evaluation and program staff then develop a written agreement-for-services which specified the type of evaluation activities to be undertaken. If funds are remaining after these efforts are accomplished, they are returned to the program. Abuse by either side is then mitigated by this agreement, and evaluation and program activities typically receive their available share of attention. This form of resource allocation seems desirable in that it allows greater autonomy--avoiding the reliance upon high administrative support (as in site B) and circumvents the constraints imposed by local officials who are less favorably disposed towards evaluation (as in Site G). We recognize that Congress cannot legislate issues of independence at the local level, however, promoting options for altering allocation procedures to facilitate independence seems warranted.

Policy on Independence. Few school districts have explicit policy regarding independence of the evaluation unit. The Austin Independent School District is a remarkable exception. District Superintendent, Jack L. Davidson, for example, has proposed that to assure independence, the school administration must support the idea through explicit policy and administrative action. The Austin policy makes the unit responsible to superintendent rather than to specific program directors, encourages open communication between the evaluation unit and the public, media, and school board, and prescribes operating guidelines for assuring the unit's independence. The latter include vesting responsibility for hiring evaluation staff, designing evaluations, and managing all federal evaluation funds to the evaluation unit. Reports issued by the unit carry only the names of unit authors and program staff are forbidden to alter such reports but are encouraged to respond to and comment on reports. The Austin procedures are summarized in the exhibit which follows.

Credibility and Trust at the Local Level. Judging from our round-table discussion with school board members, there is substantial trust in the established evaluator. The trust is necessary we believe since school boards normally have little expertise in technical aspects of evaluation and few resources for independent audit.

Further, our round-tables on utilization and survey suggests that credibility is built up gradually. For instance, Mesa School District's James DeGracie stresses the point, arguing that it is only after answering questions repeatedly that trust accrues and is justified.

Recent surveys of 39 large school district evaluation units, by Austin's Paula Matuszek and Ann Lee, appear to support the idea that established units do have high credibility. Few of the directors, for instance, maintain that establishing credibility is a good justification for spending money on independent assessment of their unit. The improvement of the unit's evaluation designs and communication are far more important justifications. It is clear, however, that these directors of large school district evaluation units believe that independent external evaluation of their units will have incidental but notable positive effects in the sense of increasing prestige, credibility, and the like.

EXHIBIT A

Independence of Evaluations by Office of Research and Evaluation

The Office of Research and Evaluation has the independence necessary to assure unbiased, forthright reports of the district and of individual programs' standing or achievement. This independence is insured by the following procedures:

1. Evaluation staff members are selected by and responsible to the Office of Research and Evaluation.

2. Final authority with respect to evaluation designs until they are presented to the cabinet and board rests with the Office of Research and Evaluation.

3. Funds for evaluation are administered by the Office of Research and Evaluation.

4. Reports shall bear only the names of evaluation office staff on the title page. A separate sheet listing program staff members may, however, be inserted at another point in the report.

5. Reports prepared by the Office of Research and Evaluation are not altered in any way by AISD staff members who are affected by the evaluation although all members shall have the right to respond or object to any evaluation report and to have their comments presented along with the report.

From: Davidson, J.L. The research and evaluation unit: Helping your school board make decisions. Paper presented at the Annual Meeting of the American Association of School Administrators, Atlantic City, New Jersey, 1976. (Available from the author, Superintendent, Austin Independent School District, Austin, Texas).

Apart from the natural interactions which engender trust between evaluator and audiences for evaluation, specific administrative action is warranted. For Austin School District's Jack Davidson, the action should take the form of assuring independence, quality and credibility. In particular, he maintains that administrators must be committed to the free flow of both positive and negative information and the superintendent's responsibility to bring the school board, the public, and the staff to this philosophy. His remarks before the American Association of School Administrators make plain the need for open reporting and independence of the Evaluation unit.

Candor in Reporting at the Local Level. The pressures affecting candor in local reporting to federal agencies are peculiar. There is an inclination to report honestly, but there are also reservations because of a fear that funds will be cut if the report comes out unfavorable. On the other hand, individuals recognize that funds are rarely if ever cut on the basis of evidence, and such decisions are based on other factors especially "politics." Complicating this is the fact that targets or goals can easily be set so that progress appears to be made, eliminating the possibility of grant termination on grounds of failure, and yet assuring that funding on the basis of need will continue so as to permit achievement for particular goals.

These issues of integrity and honesty also surface with regards to outside contractors, when they are used. One reason to hire an individual external to the district office is to ensure that information and results will be obtained which are more free from bias resulting from internal pressures and prejudices. Trust is placed in the objectivity and integrity of the contractor. However, at the same time, observations have been made as to the lack of research integrity of some hired contractors, either due to the district's naivete about selection, the lack of available qualified individuals, or, surprisingly, the nature of the contractual process itself. Contractors themselves have cited incidents of lost reports, district pressures (implicit or explicit), and purposive dilution of reports. Thus, it must be noted that district or program bureaucrats themselves often play a role in reducing the integrity of outside evaluations.

Some reports are undoubtedly withheld but the incidence of this from our site visits appears to be small. Unpleasant reports simply do not get wide circulation; more interestingly, they can result in harsh evaluations of the evaluator's competency in general. David's case studies of 15 school districts generally confirm this for Title I. Positive standardized test results generally serve to foster positive feelings toward the evaluator and his/her capabilities. Negative results are ignored or "explained away as inappropriate."

Constraints on Capabilities

The constraints on capabilities do not depend solely on individual skills, but also on factors associated with the evaluation context. In this section, we address those issues that inhibit both federally required evaluation efforts and the conduct of special evaluation activities.

Inadequate Numbers of Evaluation Personnel. No clearcut criteria exist for determining the number of individuals needed for adequately conducting evaluation activities at the SEA and LEA levels. The size of the staff is partly determined by the kinds of activities assigned and also by the size, number, and nature of the specific federal programs. For example, SEAs can be responsible for aggregating and reporting data to federal agencies and providing technical assistance to their local districts in submitting the proper information. While it may take only a few individuals to compile the necessary statistics for federal agencies, providing evaluation assistance to LEAs requires more personnel. Once additional activities are undertaken (e.g., the conduct of special studies), the required numbers and capabilities of SEA staff quickly multiply.

While federal evaluation reporting requirements have not subtantially diminished over the past decade, it is unclear whether the number of evaluation professionals in SEAs has increased accordingly. Evaluation staff positions are still heavily dependent upon federal funding and state allocation of these administrative monies to professional positions. For example, in State VI almost 60% of the evaluation unit's professionals were supported by federal monies, and results from our phone surveys provide further evidence of this dependence. It is unlikely that this situation will change, even in an era of increased accountability. In fact, in State IV this atmosphere has been interpreted to mean reduced government bureaucracy across all SEA departments, and the evaluation unit has lost two of its five staff positions as a result. It should be noted that this SEA serves over 1,000 districts, similar to State VI, but has one-twelfth of the staff. Consequently, it is not surprising that State IV does not engage in much technical assistance to its LEAs while State VI has created its own technical assistance unit.

The same problems plague LEAs. It was frequently noted in our site visits that the elimination of federal evaluation monies would result in the demise of evaluation units and personnel, given school district's declining financial resources. Employing Webster and Holley's criteria for evaluation unit staffing, two-thirds of the units in our site visit sample were understaffed. Decline of units has also occurred; for example, in Site J the evaluation staff has fallen from 22 to 2 full-time professionals within the last 7-8 years. Complaints as to "too few staff" were also noted by over 90% of the Directors of Research and Evaluation units in the CSE survey.

Inadequate staffing levels can paradoxically lead to underutilization of evaluation units in both LEAs and SEAs. For example, in our SEA phone surveys, we found that evaluation units did not typically handle all evaluation activities associated with federal programs. Vocational and Special Education programs frequently used their own staff for available reporting. The same situation was true for LEAs in our site visit sample. Only in those districts which were adequately staffed (by Webster and Holley's standards) was the majority of evaluation activities associated with these programs conducted by the unit.

The Expansion of Technical Assistance in Evaluation. Complying with federal evaluation reporting requirements does not require highly sophisticated evaluation capabilities. However, it does require that technical assistance in meeting these evaluation requirements be made available, especially as mandated activities increase in their complexity. For example, since the emergence of the Title I Evaluation Reporting System, the Technical Assistance Centers have been instrumental in training individuals responsible for evaluation reporting in SEAs and LEAs. This has included such activities as advice concerning the selection of program participants, the use of tests, and the choice of an evaluation model and the development of computer software. The provision of these services specifically targeted at evaluation has been crucial, given that every district receiving Title I monies must comply with these mandates and often do not have the trained personnel to do so on their own.

However, other federal programs have not enjoyed this extensive assistance, although SEAs may still be required to provide these services to their respective districts. For example, while one role of an SEA may be to provide technical assistance in Title VII evaluation, all SEAs are not adequately staffed or financed solely for this activity and thus have fewer resources upon which to draw than their Title I counterparts. At the same time, districts have voiced the need for aid in selecting competent outside contractors and answering other evaluation-related questions. Title VII, Special and Vocational Education staff in State II, while noting the need for technical assistance, were doubtful whether they had the adequate numbers of trained staff to provide this service to their LEAs in such areas as developing evaluation designs and alternative models.

Ways of Enhancing the Quality of Evaluation. The issues relating to technical assistance in the previous paragraphs have primarily focused on ensuring that capabilities exist for minimally complying with federal evaluation reporting requirements. However, emphasis needs to be directed at not only complying but also providing opportunities for improving evaluation efforts--especially in districts with sophisticated evaluation personnel.

A good example of how opportunities could be promoted is analogous to the contracts allocated under Section 183 of Title I. These grants are currently awarded to SEAs for the refinement of the Title I Evaluation Reporting System. This mechanism makes it possible for SEAs with competent staff to develop a proposal and receive funds to examine such issues of quality control, methodology, and cost-effectiveness. The same opportunities could be offered to competent LEAs across a variety of program areas.

At present, there exist few opportunities for district evaluators, competent and eager to conduct evaluation research to improve methods and examine issues related to federal education programs. As Webster and Stufflebeam have indicated, federal funds have not been targeted at facilitating LEAs to answer questions beyond those generated from required efforts. It is not unrealistic to assume that LEAs with highly competent staff can propose and conduct studies which can enhance general understanding of educational evaluation.

In addition, there are few mechanisms to facilitate staff development in evaluation for SEAs and LEAs. Title I Technical Assistance Center sponsored workshops represent one available option, but these workshops do not cross-cut other federal programs. Further, it is unclear whether one-to-three day workshops in evaluation are capable of providing more than minimal exposure to evaluation topics. Other mechanisms might include the provision of monies to promote university and LEA/SEA relationships or through the sponsorship of evaluation training fellowships. Concerning the first alternative, evaluation staff could receive both training and advice in conducting their assigned activities, and graduate students could obtain actual experience in educational evaluation. Sponsoring a position where such individuals as researchers or professors could participate in evaluation activities could improve the quality of information and methods used in SEAs and LEAs for evaluating federal programs.

Standards for Selection of Outside Contractors. Outside contractors are hired for a variety of reasons: to enhance independence, limited LEA resources, and for quality assurance audits. In terms of federal evaluation reporting requirements, they are typically hired by local school districts to collect information required by Title I and Title VII mandates. However, we encountered some local need for guidance in selecting appropriate contractors. For example, in State II where districts are required to hire outside contractors for Title I programs, guidance is provided by the SEA in alerting districts as to the types of tasks they should expect of their contractors (e.g., classroom observations). They are also made aware of their rights in the contractual arrangement and the problems which can result. This procedure was devised by the SEA Title I evaluator and the TAC.

However, this assistance is not so commonly provided in Title VII programs. When there is an evaluation unit present, their staff often can monitor the process and help to ensure that a competent individual is selected. However, this is not always the case and many LEAs receive little guidance as to what standards should be employed in hiring outside contractors. School boards and superintendents often complicate the process by only looking at the price tag for the evaluation rather than the skills of the bidder. Given that these contracts are typically very small, this does not encourage large reputable firms to participate in the bidding process. Districts should be provided with standards and guidelines to assist them in the contractor selection process and help prevent the possibility of obtaining poor quality evaluation.

In our interviews, we encountered numerous instances where outside contractors received a rather insubstantial amount of money to perform the required evaluation.* Similar, small allocations were observed in many LEAs for Title I evaluations. The major factor contributing to this situation is the absence of any standards by which to judge whether sufficient resources have been allocated. According to Freda Holley, the Austin School District Superintendent's office has issued guidelines as to the percentage of program costs which should be allocated to evaluation. The percentage set-asides vary in accordance with the size of the award: They recommend a 10% allocation for program awards under half a million dollars; a 7.5% allocation for a million dollar award and between 4.5% and 7.5% for awards over a million dollars. It's impossible for us to judge whether these guidelines are sufficient. Time did not permit intensive investigation. But we believe this problem is pervasive and serious enough to warrant further investigation into the consideration of guidelines on contract size for various levels of evaluation effort.

---

*See Chapter 4, Section 6.

Footnotes

[1] Full citations to the documents cited here and elsewhere in the chapter are given in Section 8, References. The citations are given by author named in the text. If no individuals are named, then the citation is to an organization such as the Evaluation Research Society, or to an institution such as the US General Accounting Office, that is identified in the text as the producer of the document.

[2] Parts of this section are excerpted from a review of the literature by Boruch and Wortman (1979). See the references in Section 8.

CHAPTER 6. HOW ARE THE RESULTS OF EVALUATION USED?

Robert F. Boruch, Laura Leviton,
David Cordray, and Georgine Pion
(Sections 6.1-6.7)

Laura Leviton and Robert F. Boruch
(Section 6.8)

Sensible people regard nothing as useless.

LaFontaine
Fables V. 19

Of course, every half-crazed accumulator
of refuse, who lives among old bottles,
stacked newspapers, and the like, regards
himself as eminently sensible.

Bergen Evans

6. HOW ARE THE RESULTS OF EVALUATION USED?

The argument about whether evaluations are used is justified. But it is not always well-informed and it is often confusing. To illustrate, we encountered a Congressional staffer who announced flatly at the beginning of an interview that evaluations were useless to his Committee. Five minutes later, he said that his committee used evaluations regularly in guiding the questioning of witnesses. Similarly, at Site D, we encountered a program director who told us the results were not used, only to discover later that he meant "not used by the administration," for they had indeed been used by staffers. We believe the confusion, or at least inconsistency, implied here underlies some of the conflicts among managers, evaluators, and policy makers regarding use of evaluation results. Part of the problem lies in agreement on what constitutes use of evaluation results, so we present some functional definitions in the next section. Part depends on the audience for the evaluation. Information useful to some audiences is immaterial to others. This too is considered briefly.

Part of the argument over what information is used also lies in flawed memory, gratuitous response, and self-interested response. We obtained underestimates in cases where respondents relied solely on memory and overestimates where, we believe, they reckoned that declaring information use was important. It is for this reason that we've enumerated probable sources of bias in a later section of the paper. Moreover, we have developed case histories based on corroborated evidence to assure that we can assay use and nonuse in at least a few clear instances.

6.1 DEFINITION OF USE AND AUDIENCES FOR EVALUATION RESULTS

The absence of any uniform definition for "use of evaluation results" underlies some of the argument about whether they are indeed used. The absence of definition certainly makes it difficult to verify claims of use, and to decide how evaluation budgets should be set. A number of efforts to assess the use of research and evaluation results have been undertaken over the past four years. These efforts, supported by NSF and NIE among academic researchers and by GAO within its organization, have resulted in a clearer picture of the use of evaluations. The three broad functional definitions presented here are based on that work.[1]

Use of Information in Making Specific Decisions

This use may involve modifying program operations or regulations, developing legislation, or constructing specific policy. The NIE Compensatory Education Study, for example, clearly influenced the form of amendments for Title I programs and some program operations. The evaluation of Follow Through did not lead clearly to major modifications of the program nor to any major specifiable decision in the legislative arena, though findings were used elsewhere. The use of evaluations to make specific decisions is not a frequent event, but it is not uncommon either. The case studies presented later illustrate how the process occurs in instances we have been able to verify. Pockets of routine use in making specific decisions, such as use by the Joint Dissemination Review Panel, are also considered later.

Use of Information to Enhance Understanding of Issues

This use encompasses understanding issues, providing context and background for policy development, and influencing ideas and attitude about a program. For instance, the House Report on the Education Amendments of 1978, notes that prior to the NIE Compensatory Education Study, evaluations had led us to believe that compensatory education was difficult to achieve and indeed not succeeding terribly well. The Report noted that the pessimistic view has been changed to one of greater optimism by the NIE report, and the consequences of such optimism may be far reaching. This also implies of course that earlier reports were also used in the simple sense of supporting a view that the program was not a splendid success in altering the intellective achievement of children.

Use of Information to Persuade Others or to Confirm One's Beliefs

The use of information to persuade others, to argue for program changes and levels of program support, and other related uses of evaluations are common. For example, former DHEW Secretary Califano cited evaluation evidence in testimony supporting particular program changes in Title I, during reauthorization hearings. Almost all of the witnesses representing nine states and testifying in March 1979 hearings of the House Subcommittee on Elementary, Secondary, and Vocational Education cited positive outcomes of local and state evaluations to argue for funding for Title I programs. This use of evidence includes supporting or confirming one's own beliefs.

Using information for rhetorical purposes is legitimate relative to some standards. They are clearly not legitimate relative to others. Results of badly designed evaluations, for instance, have been used to argue for reduced budgets and for increased budgets, for modifying regulations, and for keeping evaluations as they stand. Well-designed evaluations may lead to less equivocal conclusions, but the recommendations drawn may have little or nothing to do with the data.

Different Strokes for Different Folks

Audiences for evaluation results have been considered in Chapter 2. To summarize here, they include policy-makers, managers, and oversight agencies at national, state and local levels. At the local level, the audiences can also include parents, parent advisory groups, and teachers. Any particular audience contains individuals who are indifferent and others who attend to results. There is considerable variation across programs, across school districts, and across states. Focusing on particular audiences during the evaluation planning process is critical simply because the information made available to one audience may be perfectly useless to another. Finally, it can take a good deal of time to decide what information is most useful to which audience if we may judge from the 18 months required to set up evaluation and reporting plans by the Bureau of Education for the Handicapped, and over six months planning time required by the NIE Compensatory Education Study.

To illustrate the last point, consider Henry Brickell's examination of the kinds of information demanded by NIE, OE, ASPE, OMB, GAO and others, of a NIE-supported evaluation of a career education project. NIE and OE wanted answers to a variety of questions about program execution and about the effect of the program on children. BOAE, the management group within OE that would be the most direct consumer of the evaluation, was most interested in three questions: What can BOAE learn from this evaluation about running career education programs? Is the evaluated program transportable to other sites? What can this evaluation tell us about how to implement the program? ASPE was primarily interested in whether this program evaluation reflected whether NIE was doing work in the areas in which Congress and the public were interested, and whether the focus of NIE should change. OMB was not interested in the evaluation report per se, but in the rationality of NIE's program plan. GAO was interested in any information needed for audits, but especially effectiveness information. State legislators interviewed about their information needs requested every single kind of report available.

The managers of the evaluated programs and the six local education agencies that had adopted the program had other concerns. The managers wanted to know how to improve the operating program and how to inform the public about the project. Program specialists concentrated on reports dealing with their area of specialization, in efforts to improve the curriculum.

School board members in districts that might adopt the program wanted to know how the project was introduced into the six LEAs that had adopted the program originally. This information would assist them in understanding the feasibility of operating the program and the degree to which it would be acceptable to the community. Superintendents also wanted information on the introduction of the program into new LEAs, but in addition, reports on development of new tests for the program, and a new teacher manual were also in demand. Professional associations were interested in parent, pupil and teacher attitudes toward the new program, and in the field tests of curriculum materials.

UCLA's Alkin, Daillak, and White's intensive five case studies of five local education agencies reach analogous conclusions. Considering the use of results of evaluation in a Title IV-C program, a Title I program, and others, Alkin et al reiterate the importance of the different audiences for evaluations. States use evaluations quite differently from school districts and frequently local uses had little to do with the reporting requirements of the states. Many of the uses made of evaluation information were determined jointly by the content of the evaluation, the situation, other information, and user characteristics. In no instance could evaluation be said to be the single datum on which a decision was made. However, the case studies do provide evidence that evaluation played a detectable role in changing thinking and in making decisions.

More generally, an analysis of UCLA data on over 200 research and evaluation unit directors in school districts, conducted by Resnick, O'Reilly and Majchrzak at UCLA's Center for the Study of Evaluation, help to confirm these results and give a more general picture. Evaluation directors said there is greater use of evaluations by superintendents and school boards when the evaluation gave information about allocation of resources. Program directors, on the other hand, were viewed more often as using evaluations when they dealt with curriculum selection and modification. Superintendents and teachers also tended to use evaluations more for this latter purpose.

6.2 SOURCES OF DATA ON THE USES OF EVALUATION RESULTS

Since 1977, several major studies of use at local, state, and federal levels of government have been undertaken. For the local level, these include SRI International's Study of Local Uses of Title I Evaluation and their Evaluation of the National Diffusion Network, the UCLA Center for the Study of Evaluation's survey of school district evaluation units and Alkin's case studies for school district uses of evaluation, Rand Corporation's Study of Federal Programs Supporting Educational Change and Datta's critique of Rand's report, and the Huron Institute's current Study of Local Uses of Evaluation. Local school district offices of evaluation, such as Austin's, have undertaken smaller studies, which have been reported in the professional journals and they are no less useful.[2]

Investigations of activity at the state level are less frequent. They include SRI International's study of the National Diffusion Network, the Mitchell study of utilization by state legislators, Joan Bissell's work on uses by the California state legislature, and the Hope Associates study of Title I Technical Assistance Centers.[3]

For the federal level, we rely on the Office of Education's Annual Evaluation Report and related reports, Hearings on Costs, Benefits, and Utilization of Evaluation by the U.S. Senate Committee on Human Resources, and pertinent reports of other committees. We also use results of surveys of Congressional staff members conduted by David Florio, Harrison Fox, and Hillel Weinberg, the last two being current and former Congressional aides. In addition, we include information from selected case studies, developed by NIE staff members: Datta's study of Headstart and Millsap's case study of use of evaluations in regulation writing. Finally, we rely on our case studies and surveys. Details are given in Appendix 3 and Footnote 4 of this chapter.

6.3 USE OF EVALUATION RESULTS AT THE NATIONAL LEVEL

The Annual Evaluation Report

The Office of Evaluation and Dissemination has issued a formal Annual Evaluation Report on programs administered by the U.S. Office of Education since 1971. Information on uses of evaluation reports has been routinely reported since 1974. Early reports confined attention to uses of evaluation "studies" while the most recent cover "evaluation activities" including the generation and distribution of manuals for local school district use. There is notable overlap in the studies cited from year to year, with new ones being added as they are used and old ones eliminated as a "use" becomes obsolete. The latter includes, for instance, use of studies on alternative formula for Title I formula allocation for Public Law 93-380 in 1974-75.

In the Annual Evaluation Report for 1979, some 42 evaluation activities are enumerated in the section on use of evaluation products.

. 6 items refer to production of technical manuals

. 3 items refer to production of data tapes

. 33 refer to reports on evaluations.

The presentation is summarized in the exhibit attached. The index of use for manuals is distribution and sales. So for instance some 14,000 Handbooks for guiding LEA efforts to evaluate the impact of their programs have been sold. The index of use for tapes is distribution. For instance, a survey of state and local use of education funds, required by law and conducted through OED, is said to have been distributed to a variety of government agencies including the Congressional Research Service, universities, research institutes, and the National Education Association.

The 33 evaluation reports mentioned in the Annual Evaluation Report can be classified only very roughly into categories to obtain some feel for the products.

Exploratory, planning and needs assessment	10
Process, implementation, compliance	24
Estimating effect of programs on clients	6

The total does not add to 33 since some studies, of bilingual education for instance, have multiple objectives.

Of the 24 evaluations bearing on implementation of programs, almost all are said to have been used in management decisions of one kind or another. So for instance, evaluation of state plans for career education programs led to half the states revising their plans. Reports on higher education were reported to have been used in developing budgets. Regulations were changed at least partly because of evaluations of the Emergency School Assistance Act, Title III, Desegregation Assistance under Title IV, accreditation practices, and Exemplary Vocational Education Programs. Some led to changes in internal management procedures, e.g., evaluations of Title I migrant education record systems, of earlier evaluations of state programs under Title I, of operations under the Emergency School Assistance Act. No more than a half dozen of the implementation studies appear to have been used in amending law or in authorization decisions: Emergency School Assistance Act, Sustaining Effects Study, accreditation, ESEA Title IV, Compensatory Reading, and Vocational Education for the Handicapped.

Of the half dozen or so evaluations which address the question about what the effects of services are, the uses are mixed. The bilingual study, resulted in changes in law, policy, and regulation and has had some effect on appropriations. Review of evaluations of career education programs resulted in management decisions to approve some programs for dissemination. Evaluation of Follow Through was cited specifically in hearings, and some

Exhibit for Chapter 6

FY 1979 Annual Evaluation Report
Uses of Evaluation Activities

	Topic or Area	Product	Evidence on Use
1.	Title I Models	Evaluation Models	.percent of school districts using models
2.	Title I	Data	.Public Testimony
3.	Title VII Bilingual	Report (Effect)	.Modification in Law .Internal audit and tracking created .Management change .Report (HR 7555) citation .Regulations
4.	Annual data	Tapes	.Users, such as Rand, UC, NEA
5.	Title III Special Projects: NDN	Identification of 127 Effective Projects	.2,185 adoptions during 1978-79 .Rhetorical emphasis on implementation
6.	Career Education Search	Exemplars: 7 out of 257 with evidence Local need for evaluation (Effect)	.JDRP/NDN management
7.	Career Education	Evaluation Handbook	Commercial publication
8.	Career Education	"Evaluations" of State Plans for Career Ed	Half of states revised plans
9.	Methods	.Handbook: measuring project impact .Guide to validating gains .Sampling procedures	14,000 sold 12,000 sold no information
10.	Bilingual Materials Development	.Study of state of materials .Inventory of materials .Proposed regulations	proposed regulations

12.	Higher Education Title III	.Report on institutions	Changes in regulations, but unspecified
13.	National Longitudinal Study of the High School Class of 1972	.Data analysis	Use in *Bakke* case
14.	Consumer Protection	.*Handbook*: Student's consumer guide to college and occupational ed	"Commercially distributed"
15.	Higher Education	*Technical Report*: Alternative approach to distribution of aid to students	No evidence on use of distribution
16.	Higher Education	Report on forming state Guarantee Agencies	Presented in testimony by Commissioner
17.	Higher Education	Reports on higher ed financing	."Used by staff in developing budget" ."Used by congressional staff in reauthorization."
18.	Emergency School Assistance Act	Study of Nonprofit Organizations (NPOs)	.Specific changes in draft regulation
19.	Emergency School Assistance Act	.Study on need and management	.Specific changes in law .Specific management changes
20.	Desegregation	.Handbook for Integrated Schooling	70,000 copies "distributed"
21.	Desegregation Assistance Title IV	Study	Substantial revision of regulations
22.	ISAA	Sustaining Effects Study	Used in reauthorization (*HR* 15, *Hearings*)
23.	Follow Through	Abt Study	.Citation in Report 95-1151 on (HR 7577) .JDRP Review of 21 projects
24.	Basic Skills	*Conference*	.Feedback on evaluation design unspecified

25.	Budget	Budget Projection Models	.No specific evidence on uses cited.
26.	Loans	Survey of lenders	No specific information
27.	Accreditation	Study	.FTC regulations change .1976 law change .Carnegie Commission Report
28.	College Finances	Study	.Use in policy paper
29.	ESEA, Title IV	Study	.Nonspecific influence on regs
30.	Educational Change	Study	.Nonspecific Congressional action
31.	ESEA Title I Neglected Children	Study	."Report shared"
32.	ESEA Title IV	Study	Specific Citation in Hearings
33.	Compensatory Reading	Study	Specific Citation in Hearings
34.	Upward Bound	Study	Regulations revised
35.	Planned Variations	Report	Letters attesting to usefulness
36.	TV	Report	Book Award
37.	Voc Ed/Handicapped	Report	1976 Amendments change
38.	Voc Ed/Disadvantaged	Report	"made available"
39.	Exemplary Voc/Ed Programs		Changes in regulations
40.	Community Right to Read	Report	"Significant contribution to establishing guidelines for Reading Academies
41.	Title I Migrant	Report	Conversion for fund allocation
42.	Reanalysis of ESEA Title I Reports	Report	Executive document (unspecified)

management decisions were made in the sense that some Follow Through models were reviewed as to effectiveness and approved for dissemination by JDRP. Results of Title I assessments were used theoretically, to argue for continuation or expansion of the program. Evaluation of Upward Bound programs resulted in some changes in regulations.

This summary and the information provided in the Annual Report is a bit too terse to do justice to the topic of management uses of information. For instance, the evaluation of the Experience Based Career Education Program (ECBE) developed by NIE, was clearly used in the formulation of regulations for exemplary projects under Part D of the Vocational Education Act. The evaluation suggested that this experimental program was notably successful in achieving desirable goals. As a result, the regulations gave priority for funding under Part D to replications of the ECBE program. Other proposals for projects under Part D would have to show that they were at least as effective as ECBE, through evaluation evidence presented to the Joint Dissemination Review Panel. Sixty percent of subsequent grants under Part D were replications of ECBE, while another 25% combined ECBE with other programs. OE has used evaluations to improve its National Diffusion Network. For example, a study of the Project Information Packages showed that, although PIP's brought about improvements, personal assistance was also necessary to implement innovations in LEA's. As a result, developers were funded both to develop materials and to provide personal assistance in implementing innovations. Several evaluations (RAND change agent study, evaluation of PIP dissemination and implementation) indicated the need for assistance to LEA's in tailoring the innovation to their specific needs, and developers were funded to do so.[5]

We have no evidence that the catalog of uses of products enumerated in Annual Evaluation Reports is untrustworthy. The information presented, however, is often not sufficiently precise to permit an outsider to verify it. To be sure, there are explicit references in the 1979 Report to specific hearings and to law in some cases, five or so in disucssion of Title VII bilingual education. But there are about 17 instances in which citation is incomplete. "FTC regulations. . .developed partly on the basis of findings," "proposed regulations," and so on. Apart from this, the actual title of the evaluation report which is said to have been used fails to appear in a dozen or so cases. Congressional reports are often similarly incomplete in citation. It is on account of imperfections in citation, and in the interest of verifying OED contentions that we undertook the case studies discussed later.

The OED list is also a bit confusing in the simple sense that evaluations, technical manuals, and production of data tapes are all combined. The catalog is uninformative about periodicity in use of information. Because it is cumulative and because citation is imperfect, any given entry may refer to an evaluation completed anytime during 1974-1979. This makes the report of limited usefulness as an index of productivity of the evaluation unit.

The catalog does not enumerate studies that have had little use to policy makers, managers, and oversight agencies, and that have evoked little professional interest. It is not unreasonable to do so, in the interest of accounting for evaluation monies. The task need not be left to the agency.

Congressional Staff Views

The views of Congressional staffers with responsibilities pertinent to evaluation are mixed but tend to be positive. So for instance, David Florio conducted interviews with 26 staff members concerned with educational issues. They found that the focus of these staff members was less on general knowledge than the specific policy issue at hand. General knowledge about research and evaluation did filter up to these users from other sources. These sources, ranked in order of Congressional reliance on them, include:

GAO, CRS and other Congressional agencies	(37% of staff mentioned them)
Federal agencies	(29%)
Professional associations	(12%)
All other sources (for example, home based)	(22%)

Washington-based sources of information predominated (80%).

Evaluations must compete with other sources of information Congress receives. In the Florio survey, about a third of staff members felt that research was consistently important as a source of information. Another third felt that the importance of research information depended on the issue at hand. Most respondents felt different types of research information were useful for their purposes at different points in the legislative cycle. A majority of staff members believed research was useful for development of issues, deliberation, and oversight. However, the group was split almost evenly as to whether they felt research was useful for decision-making itself. This may reflect the Congressional view of what the decision-point is, namely, a vote and compromise. We know from studies of legislators that as votes approach, positions of legislators harden, such that research information becomes less influential.[6]

An earlier survey by Harrison Fox, a Congressional staffer for former Senator Brock, supports the finding that evaluations are a valued source of information for oversight, at least in principle. Senate staff participating in the survey believed that evaluations were effective for this purpose, ranking only behind hearings and meetings, staff communication with agency personnel, staff investigations, and audits of agency programs.

In the Florio survey, different types of research information did not have equal importance to Congressional audiences. Cost of the program to the government was most important, and student's achievement ranked second. A variety of other information followed in the rankings -- but note that costs and achievement scores represent goals on which everyone agrees, for almost all educational programs. The importance of other information undoubtedly depends on the issue at hand.

Congressional Use of Evaluations: Committee Reports

Federal agency staff members often complained about the extent to which evaluations were not or could be used by the Congress. We could not investigate this problem intensively in the time available. However, we did review various committee reports to understand the extent to which evaluations were mentioned, and the extent to which a reference to evaluation carried sufficient information for the bureaucrat or outsider to determine what evaluation was useful. The following remarks summarize some of that work. Case studies which trace use of evaluations in Compensatory Education, Bilingual Education, Innovative Programs, and others are given later.

The contents of the Report of the House Committee on Education and Labor on the Educational Amendments of 1978 reveals nearly 30 references to the NIE Study and one contractor is identified. There are 5 references to GAO assessments and 4 references to OE evaluations. Two OED contractors were cited without reference to OE. The Report of the Senate Human Resources Committee partly duplicates the language of the House report. In both Reports, criticism of OE was reiterated. The main basis for criticism was the GAO's report on OE evaluations.

The 1979 Report of the Senate Committee on Appropriations provides a terse rationale for each of the 85 or so items on which judgments are made. There are 11 references to evaluation, including 3 GAO reports. There are two verifiable references to OED supported studies. None of the references are specific in the sense of specifying a title of the study or document. The conscientious inquirer can presumably go to the Hearings.

The 1980 Report of the House Committee on Education and Labor reviewed elements of each of thirteen Titles. There are six references to evaluations, including one each produced by OE, GAO, and the IG. The only specific reference is given to the evaluation of the Fund for the Improvement of Postsecondary Education by Sol Pelavin. The office which sponsored the study, ASPE, is not acknowledged.

Reports of the House Committee on Appropriations for 1977, 1978, 1979 cite evaluations on nine occasions. Two of these are verifiable references to OE studies and one to GAO.

From this small search, we conclude from the more conscientiously constructed reports that: (a) On average, evaluation is mentioned and presumably used in 1 out of 8 or 9 cases in which judgments about individual budget items are made. In the remaining 7 or so cases, there has been no evaluation at all, or there has been a useless evaluation, or a useful evaluation has not been cited or has been ignored, or no one thought to mention it. (b) The citations to studies are often vague and, at least at times, studies which appear to have been used are not acknowledged. (c) Citation rate of OED mounted studies is not much different from citation rate of GAO studies in recent years.

Exhibit for Chapter 6

References to Evaluation in: U.S. Senate, Committee on Human Resources. Report: The Educational Amendments of 1978, 95th Congress, 2nd Session, Report 95-856, May 15, 1978.

Elementary and Secondary Education: Title I

A. Grants for the Educationally Deprived: 11 major references to the NIE Compensatory Education Study, one to an unspecified OE Study (probably sustaining effects), one to a GAO study.

B. Programs operated by state agencies

1. Handicapped: No reference*
2. Neglected and delinquent: Nonspecific reference to an evaluation (actually OED)
3. State migrant program: Nonspecific reference to DHEW information.
4. Payments for state programs: No reference

C. State Administration: 3 major references to NIE Study

D. Federal Administration:

1. Evaluation: Nonspecific reference to GAO review of OE annual evaluation report
2. Complaint resolution: No reference
3. Audits: Reference to the Assistant Secretary DHEW Sanction Study
4. Withholdings: No reference
5. Policy Manual: NIE Study referenced

E. General Provisions

1. Basic Skills: Reference to NAEP
2. Metric Education: No reference
3. Arts in Education: No reference
4. Consumer Education: No reference
5. Youth Employment: No reference
6. Law-related Education: No reference

---

*Note: In this and succeeding exhibits, "No reference" means no mention is made of any evaluation. An evaluation may or may not have been done in each case.

7. Environmental Education: No reference
8. Health Education:
9. Correction Education:
10. Biomedical Sciences: Nonspecific reference to NIH, AMA studies
11. Population Education: No reference

Libraries, Learning Resources, Educational Innovation (Title IV): Nonspecific Reference to Rand Study of innovation

State Leadership (Title V)

1. Application:
2. Rulemaking:
3. Technical Assistance: Reference to DHEW Sanction Study
4. State Monitoring: Reference to GAO report complaint and resolution
5. Withholding: NIE study cited
6. Audits: DHEW Sanctions Study, NIE Study

Emergency School Assistance Act (Title VI): GAO report cited

Bilingual Education (Title VII): Absent

OE information mentioned; nonspecific reference to GAO report

Ethnic Heritage Studies (Title IX): No reference

Community Schools (Title X): Specific reference to OE evaluation

Women's Educational Equity (Title XI): No reference

Non-public Educational Assistance: Specific reference to NIE study

Impact Aid (Title II)

Thirteen items: No reference

Extension (Title III)

1. Adult Education: No reference, but fascinating passage on immigrant organizations as educators.
2. Indian Education: No reference
3. Teacher Training: No reference

General Education Provisions: Paperwork Commission cited.

Exhibit for Chapter 6

References to Evaluation in: U.S. Senate, Committee on Appropriations Report: Departments of Labor and Health, Education and Welfare Appropriations Bill, 1980, H.R. 1489. 96th Congress, First Session, (Report 96-247), July 13, 1979.

Elementary and Secondary Education

1 Grants for the Disadvantaged: No reference to any evaluation, despite Senate use of the NIE Compensatory Education Study.
2 Grants for the Disadvantaged - LEAs: No reference to any evaluation, despite Senate use of NIE.
3 Grants for the Disadvantaged - SEAs: No reference.
4 Evaluation and Studies: No reference.
5 Concentration Grants: No reference,
6 State Incentive Grants: No reference.
7 Support and Innovation: No reference.
8 Bilingual Education: No reference to the bilingual evaluation despite verified use but the statements depend on the AIR evaluation.
9 Basic Skills: Nonspecific reference to evaluation.
10 Achievement Testing Assistance: No reference.
11 Follow Through: No reference.
12 Alcohol and Drug Abuse Education: Reference to lack of evaluative information.
13 Environmental Education: No reference.
14 Educational Broadcast Facilities: No reference.
15 Ellender Fellowships: No reference.
16 Ethnic Heritage: No reference.
17 General Assistance to Virgin Islands: No reference.

Impact Aid

18 Formula Construction: No reference to evaluation, despite discussion of analyses by ASPE.

Emergency School Aid

19 General Grants: A GAO study is mentioned.
20 Special Programs & Projects: No reference.
21 Magnet Schools: There is a nonspecific reference to the Magnet School evaluation.
22 Grants to Non Profit Organizations: There is a nonspecific reference to a DHEW evaluation.
23 Educational Television & Radio: No reference.
24 Evaluation: No reference.
25 Civil Rights Training and Advisory Services: No reference.

Education for the Handicapped

26 State Grants: No reference.
27 Pre-school Incentive Grants: No reference.
28 Deaf Blind Centers: No reference.
29 Severely Handicapped Projects: No reference.
30 Early Childhood: No reference.
31 Regional Vocational Education: No reference.
32 Innovation & Development: Evaluation alluded to.
33 Media Services: No reference.
34 Special Education Manpower Development: No reference.
35 Special Studies: No reference.

Occupational, Vocational, and Adult Education

36 State Grants: No reference.
37 Program Improvement: No reference.
38 Special Programs: No reference.
39 Consumer and Homemaker: No reference.
40 State Advisory Councils: No reference.
41 State Planning: No reference.
42 Programs of National Significance: No reference.
43 Bilingual Vocational Training: No reference.
44 CETA and Vocational Education: No reference.
45 Adult Education: No reference.
46 Basic Educational Opportunity Grants: No reference.
47 Supplemental: No reference.
48 College Work-Study: No reference.
49 Direct Loans: No reference.
50 State Student Incentive Grants: No reference.

Higher and Continuing Education

51 Special Programs for Disadvantaged (TRIO): No reference.
52 Veteran's Cost of Instruction: No reference.
53 Educational Information Centers: No reference.
54 Developing Institutions: GAO evaluation mentioned.
55 Cooperative Education: No reference.
56 International Education: No reference.
57 University Community Services: No reference.
58 State Post-secondary Education Commissions: No reference.
59 Graduate Professional Opportunities: No reference.
60 Legal Training for the Disadvantaged: No reference.
61 Public Service Fellowships: No reference.
62 Mining Fellowships: Evaluation alluded to.
63 Law School Clinical Experience: No reference.
64 Construction Industry Subsidy: No reference.
65 Architectural Barrier Removal: Evaluative information requested.

Special Projects and Training

66 Youth Employment: No reference.
67 Biomedical Sciences: No reference.
68 Arts in Education: No reference.
69 Metric Education: No reference.
70 Consumer Education: No reference.
71 Gifted and Talented Children: No reference.
72 National Diffusion Network: Nonspecific reference to evaluation.
73 Educational Television Programming: Nonspecific reference to evaluation in children's television.
74 Push for Excellence: No reference.
75 Career Education Demonstrations: No reference.
76 Law related Education: No reference.
77 Women's Educational Equity: No reference.
78 Community Schools: Nonspecific allusion to evaluation.
79 Cities in Schools: No reference.
80 Career Education Incentives: No reference.
81 Gifted & Talented Children: No reference.
82 Teacher Corps: No reference.
83 Teacher Centers: No reference.
84 Planning and Evaluation: No reference.

Exhibit for Chapter 6

References to Evaluation in: U.S. House of Representatives, Committee on Education and Labor. Report: The Educational Amendments of 1980, 96th Congress, First Session, Report 96-520, October 17, 1979.

Title I: Outreach
  No reference.

Title II: College and Research Library Assistance and Library Training and Research.
  No reference.

Title III: Developing Institutions
Nonspecific reference to formula development
(OED not mentioned)

Title IV: Student Assistance

1. BEOG: No reference
2. SEOG: No reference
3. SSIG: No reference
4. Special Programs for Students from Disadvantaged Backgrounds (TRIO): No reference
5. Education Information: No reference
6. Assistance to Institutions of Higher Education: No reference
7. Low Interest Loans: Nonspecific references to IG report, OE evaluation
8. Work Study Programs: Smith cited.
9. National Direct Student Loans: No reference.
10. General Provisions: Reference to GAO report
General lack of information cited.

Title V: Teacher Corps and Teacher Training:
  No reference.

Title VI: International and Foreign Language
  No reference

Title VII: Construction
  No reference

Title VIII: Cooperative Education
  Nonspecific reference to "a study."

Title IX: Graduate Programs
  No reference

Title X: Fund for the Improvement of Postsecondary Education
  Reference to NTS Evaluation, no mention of ASPE

Title XI: Urban Grant University Program
No reference

Title XII: General Provisions
No reference

Title XIII: Community College of American Samoa
No reference

Exhibit for Chapter 6.
References to Evaluation in the House Reports on the 1979, 1978, and 1977 Hearings of the Committee on Appropriations (Reports 96-244, 95-1248, and 95-381).

Elementary and Secondary Education

1 Grants for the Disadvantaged: No reference to any evaluation despite House use of the NIE Compensatory Education Study.
2 Grants for the Disadvantaged - LEAs: No reference to any evaluation despite House use of NIE.
3 Grants for the Disadvantaged - SEAs: No reference.
4 Evaluation and Studies: No reference
5 Concentration Grants: No reference
6 State Incentive Grants: No reference.
7 Support and Innovation: No reference.
8 Bilingual Education: Reference to "recent evaluation study: in encouraging action on deficiencies in the program (1980).
9 Basic Skills: No reference.
10 Achievement Testing Assistance: No reference.
11 Follow Through: No reference.
12 Alcohol and Drug Abuse Education: No reference.
13 Environmental Education: No reference.
14 Educational Broadcast Facilities: No reference.
15 Ellender Fellowships: No reference.
16 Ethnic Heritage: No reference.
17 General Assistance to Virgin Islands: No reference.

Impact Aid

18 Formula Construction: No reference.

Emergency School Aid

19 General Grants:
20 Special Programs and Projects: Reference to a "recent study" that reexamination of the "follow-the-child" concept is necessary (1978).
21 Magnet Schools: No reference.
22 Grants to Non Profit Organizations: No reference.
23 Educational Television and Radio: No reference.
24 Evaluation: No reference.
25 Civil Rights Training and Advisory Services: No reference.

Education for the Handicapped

26 State Grants: No reference.
27 Pre-school Incentive Grants: No reference.
28 Deaf Blind Centers: No reference.
29 Severely Handicapped Projects: No reference.
30 Early Childhood: No reference.
31 Regional Vocational Education: No reference.
32 Innovation and Development: No reference.
33 Media Services: No reference.
34 Special Education Manpower Development: No reference.
35 Special Studies: Reference to BEH study showing 1/3 of hearing aids worn by public school children are malfunctioning. (1977)

Occupational, Vocational, and Adult Education

36 State Grants: No reference.
37 Program Improvement: No reference.
38 Special Programs: No reference.
39 Consumer and Homemaker: No reference.
40 State Advisory Councils: No reference.
41 State Planning: No reference.
42 Programs of National Significance: No reference.
43 Bilingual Vocational Training: Reference to "recent studies and reports" showing lower employment rates of limited English-speaking population (1979).
44 CETA and Vocational Education: No reference.
45 Adult Education: Reference to 1975 study of prevalence of functional illiteracy (1977).
46 Basic Educational Opportunity Grants: Reference to computer audits finding many inelligibles receiving grants (1979).
47 Supplemental: No reference.
48 College Work-Study: No reference.
49 Direct Loans: No reference.
50 State Student Incentive Grants: No reference.

Higher and Continuing Education

51 Special Programs for Disadvantaged (TRIO): No reference.
52 Veteran's Cost of Instruction: No reference.
53 Educational Information Centers: No reference.
54 Developing Institutions: GAO evaluation mentioned.
55 Cooperative Education: No reference.
56 International Education: No reference.
57 University Community Services: No reference.
58 State Post-secondary Education Commissions: No reference.
59 Graduate Professional Opportunities: No reference.
60 Legal Training for the Disadvantaged: No reference.
61 Public Service Fellowships: No reference.
62 Mining Fellowships: Evaluation alluded to.
63 Law School Clinical Experience: No reference.
64 Construction Industry Subsidy: No reference.
65 Architectural Barrier Removal: Reference to HEW data on cost impact on institutions if barriers are removed. In light of demonstrated need, funding by committee enhanced (1979).

Special Projects and Training

66 Youth Employment: No reference.
67 Biomedical Sciences: No reference.
68 Arts in Education: No reference.
69 Metric Education: No reference.
70 Consumer Education: No reference.
71 Gifted and Talented Children: No reference.
72 National Diffusion Network: No reference.
73 Educational Television Programming: No reference.
74 Push for Excellence: No reference.
75 Career Education Demonstrations: No reference.
76 Law Related Education: No reference.
77 Women's Educational Equity: No reference.
78 Community Schools: No reference.
79 Cities in Schools: No reference.
80 Career Education Incentives: No reference.
81 Gifted and Talented Children: No reference.
82 Teacher Corps: No reference.
83 Teacher Centers: No reference.
84 Planning and Evaluation: No reference.

6.4 STATE USES OF EVALUATIONS

Very little systematic information is available on the uses to which states put evaluative information. No surveys and very few case studies have been conducted. The material summarized here is merely illustrative.

Joan Bissell of California's Office of the Auditor General has documented use of evaluations by the California State Legislature. Although the links between the evaluation and the actions and deliberations of the legislature are difficult to verify, in most instances she presents plausible linkages, her own experience of the path the deliberations follow, and respondents' information about these linkages. She offers four examples. The state's School Improvement Program developed a rating system for quality of school programs, which showed little relation to pupil achievement in evaluation. This finding led to changes in the statutes governing allocation of resources to programs. In another evaluation, little relation was found between level of funding for compensatory programs and achievement of pupils. State policies on compensatory education were revised consistent with these findings. In a third case, a state program for Mentally Gifted Minors (MGM) was shown to exclude some high achievers on the basis of IQ and to have excessively high administrative costs. Deliberations were undertaken by the legislature to determine the future of the program in light of the Jarvis Proposition 13 amendment. Finally, Bissell offers an evaluation in which public schools' procedures for contracting with private vocational training were found to lead to excessive costs. Hearings are scheduled to determine alternate forms of financing such training.

Mitchell conducted an interview study of state legislators' use of social science. He found that legislators use social research when the information is brought to their attention early in the legislative process. When this occurs, research is used as background orienting legislators to the questions, and in the process of forming coalitions. As debate becomes more partisan and positions better defined, social research is used as political ammunition, and ceases to have much influence with those who are not already convinced. Mitchell found that social science expertise can be highly valued by state legislators, but is only one form of valued expertise among many.

Federal Requirements that Use be Specified.

The 1976 amendments to the Vocational Education Act (20 U.S.C. 2312) require that each state evaluate the effectiveness of programs every five years. The "results of these evaluations shall be used to revise the states' programs and shall be made available to the State Advisory Council." A separate amendment (20 U.S.C. 2308) requires that an annual accountability report" contain a summary of the evaluations. . .and a description of how the information from these evaluations has been or is being used by the State board to improve its programs." The three accountability reports we reviewed happened to be readily available. Neither is especially informative and one is dreadful.

The Annual Program Plan for 1981 and Accountability Report for 1979 for the State of ________ are:

> "The Agricultural and Home Economics Units...used the data to plan teacher conferences...The program reviews have been most valuable...for necessitating state staff persons and local school persons to participate in a joint discussion of the programs...(H)opefully, program improvement will result from these reviews should have improved their competencies and prepared them for future program reviews as well as make them better supervisors" (sic, p. 41-42).

Other material is of the same sort. No specific uses are given. Good grief.

The State of ________ Accountability Report for FY 1978 is rather general. The document says that programs "use the composite report in identifying priorities for...research, curriculum development, personnel development, student services, elimination of sex bias, and funding." Other references are to the report as a resource document for universities and an assortment of other agencies. No more specific information is furnished.

The third accountability report says that "as a result of the evaluations, the state agency has made the following commitments...sponsor regional workshops...and exemplary programs in guidance...working with local vocational directors at quarterly meetings to assist in local placement...awarded contracts for competency based materials in six subject areas... ." There is no attempt to link any of these actions to a specific evaluation finding.

We have been able to locate only two formal studies of the actual uses to which evaluative program reviews, required by the Vocational Education Amendments, are put. They were conducted by Charles Manning for the Napa County, California, School District in 1975 and 1976. Each involved asking district administrators what changed as a result of the program. The majority of tangible changes which were said to have come about as a result of the report involved additions or restrictions of resources and changes in administrative procedures. The most remarkable one was recognizing the need to hire a coordinator or director of evaluation and the decisions to hire one. The most common intangible uses are reported to have been changes in understanding of problems and improved communications.

6.5 LOCAL USES OF EVALUATION RESULTS

The audiences for results at the local level are analogous to those at the national level: administration, program staff and teachers, parents and school boards who may be both participants and oversight agents. As we said earlier, the audiences contain members who are indifferent to results. Different audiences demand different information.

Site Visits

Information on the character of local level uses generated from our site visits, is given in the next Exhibit. The conclusions we draw are:

- there is enormous variability in interest in evaluation and consequently, in the use of evidence.
- LEAs with strong research and evaluation units are more likely to use evaluations in any of the senses in which we've defined utilization.
- in LEAs without strong units, use depends heavily on the vigor and skill of the individual evaluator, and or local conditions.

Generally speaking, evaluations appear to attract attention if they suggest ways to improve program operations relative to some standard. Better tests are selected, opinions about particular features of the program elicited, side studies on very good performers or very poor performers carried out. The work may also be undertaken to have an independent appraisal which carried more weight in aruging for more funds.

In very few site visits did we find any serious attempts to document the uses to which evaluation has been put. The exceptions occur in states with Title I regulations bearing on use. For instance, ESEA Division of Massachusetts' Department of Education is one of several states that require reports on the uses of evaluation. In particular, the Division asks that in an Interim Evaluation Report, the LEA report "specific changes (which) have been incorporated into this year's project, based on findings and recommendations of the last year's final evaluation report." Rhode Island makes a similar demand of its Title I evaluators and a case study on uses is reported later in this chapter.

The following remarks rely on other studies to describe what is known about use of information at the local level.

Use in Modifying Programs

UCLA's Center for Study of Evaluation reports that nearly half of the Directors of Evaluation Units in 350 large school districts say they spend some time "modify(ing) programs using evaluation results." The time investment in this activity is not high, however. Only about 15% rank it #1, 2, or 3 with respect to time demand. Other activities, such as assessing results or worth of a program (relative to unspecified standards), account for much more of their time. Presumably the uses to which such assessments are put include modifying the program.

Exhibit for Chapter 6

Sketches on the Use of Evaluation at the Local Level from Site Visits

Site A.

An outside contractor does the evaluation of bilingual education, examining match between goals and activity, and outcomes. The report is circulated but no use was cited. The evaluation report by the outside contractor for Title I is generally completed after the new school year has begun so it is useful only to confirm performance in the preceding year, not for changes in the current year. There is no strong tradition of supporting evaluations or using them. Testimonials appearing in the school newsletter on federal or state supported programs generally get the greatest attention.

Site C.

The general emphasis of the LEA is on meeting reporting requirements. There is no emphasis on evaluation and, indeed, the major and minor decisions are based on outcomes of bureaucratic and political battles. Government requirements do not produce data which is useful to them. Evaluations initiated internally are minimal and use is made by program staff but not by administration, oversight, or policy groups at the local level.

Site D.

Instances of use were frequently mentioned, especially for locally initiated evaluation activity, but there is very little documentation available on use. Raw data required by the federal government is not especially useful by itself. Reports are made available to a variety of audiences routinely--faculty, staff, advisory groups, administration. Staff members have initiated evaluations on free time. Evaluators say that good estimates of effect of Title I programs are impossible so they focus on implementation matters. Outside contractor for evaluations in bilingual has been helpful in identifying salient questions and clarifying issues. Most of the evaluative work is done by the LEA research unit.

Site E.

A vigorous research and evaluation unit is responsible for answering questions initiated by Title I staff, school board, and administration, evaluation staff as well as meeting federal requirements. The latter are regarded as minimal but compatible with interests. Questions addressed focus on program operations in the interest of program modification, program planning, budget cuts, and also address "interesting questions" when resources permit. Evaluation is seen as useful but lack of funds is regarded as a crucial impediment. Some attempts to estimate effects of programs on clients, cost benefit analysis of alternatives, etc. are undertaken.

Site F.

A small evaluation unit whose coverage includes Title I, bilingual, and special education. Evaluation goes beyond federal requirements in the first two programs. The results of monitoring are used to modify programs and to argue for continued support, for public relations purposes. But amount of use claimed by the administration and by the evaluation department assistant are far different. The latter complained consistently that recommendations were not taken and seemed to be ignored. The evaluation unit reports directly to an Assistant Superintendent for Administrative Services.

Site G.

The research unit is responsible for Title I evaluations and these are widely circulated, readily accessible, and appear to generate reactions among parents, staff, and administrators. The unit used to spend a lot of time on evaluations leading to program modifications and now that the program has stabilized they spend most time on meeting minimum requirements. Evaluation is one piece of many entering the decision process. Evaluation in remaining programs is informal. No documentation on use of evaluations is readily available.

Site J.

Site J has a two member research unit which merely augments evaluations undertaken by staff. Use of evaluation by program developer with Title IV-C funds is remarkable: field tests lead to decisions to replicate program and tests in other sites, to dissemination of program, etc. Evaluation in voc-ed is informal directed toward filling local needs and does appear to be used. The evaluator for Title I reports regularly to parents, program administrators, and staff but most sensitive or controversial findings are discussed in private with individual staffers. He does believe some of these lead to changes in program operations. The evaluations of the Title I program six years ago lead to approval by JDRP and the program itself became the basis for an ESAA program partly on account of this evidence.

Site K.

There is an instructional research unit but program staff do their own evaluations. Evaluations in all programs stress needs assessment. Outcome assessments are stressed in bilingual and Title I. They go beyond federal requirements in some programs and use information for program modification and planning, but lack of time and resources prevent more than small incremental troubleshooting. Reports by an outside contracted evaluator of bilingual criticized the program and were used to change operations and/or staff.

Site L.

Any evaluation undertaken is relatively informal. A principal may look at tests scores to determine "impact" of Title I and report this to the school board. Head counts are a typical form of monitoring. There is little emphasis on any formal evaluation of programs, federal or otherwise. There are self-evaluations in which teams are created to examine a variety of internal activities, some federal, some not. These are undertaken every five years, but no information on use of results was available.

The major recipients of evaluation unit services are teachers and principals. On average, 32% of unit time is spent in efforts to service them. The next largest groups serviced are superintendents and board members (20%) and central office staff (17%). The average amount of time dedicated to servicing federal needs and state needs is about the same, 10% or so in each case.

According to 70% of respondents, the most consistent users of information are program directors. A bit over half reported that principals, and superintendents, and central office staff were consistent users. Very few (8.5%) said parents used information consistently and a bit more said their school board used it consistently. Uses among these audiences are said to be occasional.

Alkin, Daillak and White at UCLA reported evaluation utilization in five case studies of the utilization of evaluations. These were evaluations of: (1) a program under ESEA Title IV-C to improve retention of disruptive students, (2) a teaching approach developed through an R & D center which had been adopted as part of a compensatory education program, (3) a new component of a Title I kindergarten program, (4) a career education program under Title IV-C, and (5) a bilingual program under Title IV-C. Each case study described the community and setting of the program, the program itself, the program and evaluation administrators, the evaluation, and the types of uses to which it was put. In all five cases, at least some use was made of the evaluation information to modify programs or to change views about the program. In some of the cases, these uses were major, such as the decision by the state of California to disseminate bilingual program to other districts.

Title I Programs and Test Scores

A recent study conducted by SRI International investigated local uses of Title I evaluations required from local districts receiving federal program support. SRI based conclusions on case studies of 15 school districts. Their general conclusion was that local districts rarely use evaluative data generated primarily for federal reporting (David, 1978).

In particular, SRI found that test scores required under Title I "do not serve primarily as a means of judging the program. Local skills related tests and personal observations almost always carry more weight than test scores, and goals other than achievement are often of more interest. This finding is corroborated for test scores, genreally in an independent series of 18 case studies by Sproull and Zubrow. Test scores were one of the least-valued sources of information available to central office school district administrators.[7]

But test scores did provide "gross" indicators of program effectiveness according to interviewees in the SRI study. Again, this finding is supported by the case studies of Sproull and Zubrow. However, the SRI case studies indicate that changes in programs are marginal in any case, so that there is little opportunity to use the data in making local decisions.

Test scores in the SRI study were also used to select students into the program, an idea endoursed by most in the site visits, despite the claim that they are not pertinent for evaluating program performance.

Finally, test scores are used to confirm impressions and beliefs, so long as results are positive, according to teachers. If results are negative, according to the SRI study, they are taken as an annoyance which must be explained away. The attitudes toward tests include:

"The main purpose of test scores is to support your own beliefs. (Teacher)

"I look at test scores mainly to confirm my impressions." (Teacher)

Our own site visits also suggest that test scores alone are not and should not be a basis for major local decisions about a program. Most decisions involve small modifications in any case. The information used in small modification includes test scores, incidental observations, and attitude survey, teacher or parent criticism, and others. Test scores are, however, sufficiently important to concern program managers and teachers when the scores fail to behave as someone believes they should.

Routine Use in JDRP-NDN

The National Diffusion Network (NDN), supported since 1974 by the U.S. Office of Education, makes information on innovations available to LEAs and SEAs, innovations for which sufficient evidence on effectiveness is available. A particular innovation is approved for distribution to other LEAs in the network only if it passes muster on evidentiary and other grounds with the Joint Dissemination Review Panel (JDRP). This panel includes members of staff of USOE and NIE.

In principle, the Joint Dissemination Review Panel is a quality control device and routinizes the use of evaluation findings. The Panel meets regularly to review new programs and the evidence for their effectiveness. It has produced a manual to describe the kinds of evaluative evidence which are most acceptable. This Ideabook is readily available.[8]

Between 1973 and 1976, the JDRP "met more than 60 times, reviewed over 300 submissions, of which approximately 55% were approved." All projects which are approved become eligible for diffusion under NDN and become eligible to receive Title IV-C funds to assist the project developer in diffusion and the school district in adoption. In principle then, the award of a grant hinges on the evaluative evidence used by the Joint Dissemination Review Panel.

Most federal agency staff interviewed by Northwestern's Project staff agreed that the JDRP is good in principle. It routinizes use of evaluation, links use to financial decisions, informs the program developer about standards of evidence, and serves as a useful screening device. In actual operation, it appears to be not quite so good.

Respondents, although supportive of the JDRP and emphatic about the need for such a quality control mechanism, indicated several problems with the JDRP in practice. The first of these was the role played by panel members' judgments of "educational significance." This criterion played an equal role, along with the evaluation evidence, in the decision to fund and disseminate a project. "Educational signifance" is a concept that is necessarily less clear than hard data that support the effectiveness of a program. As a result, there may be more room for the values of the panel members and their own political needs to play a role.

Second, several informants mentioned that submitters did not uniformly define the effective elements of their innovations, but rather evaluated a whole "package" that might contain several elements. Any particular element of a program might never, itself, be evaluated, and yet adopting sites might choose this element out of all others, to adopt. OED is currently evaluating improved methods of ensuring that all such elements of innovations are evaluated before dissemination.

Adopting sites are not, at present, uniformly evaluated themselves. Such evaluation is important because the generalizability of the innovations is not known. Moreover, implementation at a new site may not be adequate. The American Institutes for Research evaluation of the National Diffusion Network found that most innovations were modified at the new site. More work appears to be needed to determine the extent to which projects can be modified before they must be called a different project. OED is currently evaluating alternative methods of characterizing the crucial elements of such projects.

Considering educational innovations more generally, however, local and state education agencies use evaluation results in the following sense. All state facilitators in the National Diffusion Network rely on review of evaluation evidence by the Joint Dissemination and Review Panel in their work. The reliance is automatic since any program they disseminate must pass a review based partly on evidence before it is eligible for dissemination.

6.6 UNDERSTATEMENT AND OVERSTATEMENT OF USE

Most federal agencies and Congressional organizations acknowledge evaluations done by others in some degree. The case studies in this chapter, for instance, rely heavily on documented references to specific evaluations and corroborated testimony.

But judgments about how often evaluation studies are used at the federal, state, or local levels are doubtless less accurate than they should be because so many uses are invisible and are not publicly acknowledged. These uses are evident in letters and memoranda which are not widely circulated or not public. They are implicit in public documents which do not specify the actual study completely and in informal communications about an education program.

At the federal level for example, a response from the Comptroller General to a legislator's question may refer to evaluations conducted by agencies other than GAO. In particular, a recent letter from Comptroller General Staats to Senators Russell Long and Robert Packwood regarding federal regulations on day care suggested that the National Day Care Center Study be used in judging the operation of the program. The Study results were indeed used later, in changing regulations and in reconciling a ten-year argument between legislators and bureaucrats. Despite the fact that such letters are public under GAO rules, and are reproduced and disseminated, commentators on use of evaluation are often unaware of them. Lacking information about such memoranda means claims about the extent of use will be understated.

Our interviews with one of three Congressional Budget Office members with responsibility for educational programs uncovered a similar phonomenon. We were told that letters written in response to a legislator's question do refer to evaluations where they are relevant. CBO does not reproduce and disseminate letters of this sort as GAO does, however. A similar obstacle to tracing use of information stems from the Congressional Research Service practice of developing confidential memos and directed writing in response to legislative inquiry. These are rarely, if ever, made public. Yet some are likely to concern education programs and their evaluation. We have not had the time in this Project to peruse nonpublic written communications by either CBO or CRS, and have not done so. The point is that estimates of information use are likely [illegible] be biased downward unless one can employ nonpublic memos.

The problem of obtaining accurate estimates of the incidence of use is more difficult with unexpected demands for information. In 1979, for instance, a TV broadcast of 60 Minutes stressed the severe problems that a school district in Michigan encountered in implementing Public Law 94-142, requiring access to education for the handicapped. The district had evidently misunderstood the intent of the legislation, interpreting it to mean that handicapped children had to be mainstreamed, and the broadcast staff itself were no more knowledgeable. The evaluation staff of the Bureau of Education for the Handicapped anticipated questions from Congress being provoked by the show. They summarized pertinent results of an evaluation study then underway and conducted a fast informal survey to establish that indeed the misinterpretation was not common nor were some of the problems prevalent. The speed of this reaction is not especially common among agencies. But is rarely documented publicly and this lack of documentation adds another obstacle to making fair estimates of the use of information.

A second broad class of invisible uses concerns the cumulative nature of the information one may obtain about programs. The single most important obstacle to making such use visible is the failure in a public forum to cite an evaluation when it did influence decisions. So, for instance, 1978-79 evaluations of needs, resources, and the like, by the Bureau of Education for the Handicapped do not cite evaluative studies undertaken by the GAO on related issues in 1974 and 1976. For the public, the Congress, and the evaluation community these may then appear to be irrelevant when they are not. Similarly, evaluations undertaken by GAO sometimes do not cite earlier studies supported by federal agencies or do not cite them thoroughly enough to permit determination of which agencies did a decent job. The main point is that the acknowledgement is not always routine and this makes some evaluations appear to be less pertinent and less useful than they actually are.

In reviewing reports from the Congress itself, we find some interest in giving credit where it is due. But the lack of uniformity in citation makes tracking utility of reports difficult. Of 14 references to evaluations in a recent Senate committee Report on appropriations for instance, a clear effort was made to acknowledge especially useful studies, and to provide rationale for decisions. None cite the particular evaluation in a way which permits one to go directly to the source or to recognize the specific source. "A DHEW evaluation" is mentioned to describe why an appropriation is made, for instance, with no other specification. Moreover, three evaluations which are clearly used in other committees are not mentioned at all. This inconsistency may be unavoidable on account of the time pressures and the priority which other matters must take. But it does make use of evaluations difficult to track, and it provides confusing signals to federal bureaucrats who are interested in legislative uses of evaluation.

Informal, unwritten communications about an evaluation are not uncommon. But they are difficult or impossible to exploit retrospectively as indicators of the use of evaluation results. So, for instance, interviews with two CRS staff members confirm that when they have questions about a particular topic they simply phone staff in an executive agency, including OED, for the answer or a lead to the answer. Two members of the Congressional Budget Office confirmed reliance on telephone conversations with OED staff as well and a third preferred other source. Similar events occur at the local level. Informal and even formal exchanges of information between evaluator and central office staff, program director, and so on are often not documented, counted, or otherwise recorded. The absence of this sort of information makes utilization difficult to track and to estimate. It does suggest that relying solely on documented uses is misleading.

A related issue is that there are uses which according to one Congressional staffer "percolate up through levels of consciousness." A small evaluation may be done by an academic researcher, be picked up by the popular press, recognized and incorporated into a legislator's or executive's thinking and ultimately into a decision. This sort of phenomena may be the most common of "uses." But it is far too subtle to examine in this report.

To illustrate how judgments about the importance of evaluations may be misleading, consider Florio's recent survey of Congressional staff. Asked to rank the significance of various sources of information, staffers suggested that Fducational R & D fell above media, polls, but well below parents, local education agencies, and professional associations. One major peculiarity here is that a good deal of R & D relies on parents and LEAs to begin with. Similarly, though professional organizations are ranked higher than R & D, those organizations must obtain their information somewhere and R & D is a source, at least at times.

Either overstatement or understatement of the frequency with which evaluations are used may emerge from interviews. Overstatement occurs. At the local level, we encountered school superintendents who suggested they personally used the information but who knew virtually nothing about the evaluations generated by their staff. We expect that at least some respondents felt compelled to say yes simply because they believed it was desirable to say yes. We also encountered respondents who early in an interview told us that evaluation was not used and who later told us it was. The change usually came about when respondents simply talked through what happens during an evaluation or what happens to a report, and identified groups or individuals to whom the evaluation was pertinent.

It is not unreasonable to expect some overstatement of utility from evaluators. Because the references to use are often not specified completely, permitting little corroboration, it is also possible. So for instance in the Annual Evaluation Report for 1979 we found references to 33 evaluation reports which were said to have been used. In 17 references to use "in proposed regulations," "in Congressional action, " and the like, the reference was not sufficiently specific to track. On the other hand, some simple uses were left out entirely: if Congress demands a count and it is supplied, whether or not Congress takes action is immaterial if the information is valid. The problem of estimating incidence of use from agency reports such as this is no less difficult than the problem encountered in examining Congressional Committee reports. The reasons for imperfect citation are similar: little time for doing so, imperfect information on use, and no formal system for tracking use.

6.7 FACTORS INFLUENCING USE OF EVALUATIONS

Five general factors appear to be important in use of information, including evaluations: Timeliness, relevance, credibility, interest, and interpretability. The task of anticipating decisions is important but far less clear.

In the abstract, these factors aren't especially helpful. One of our interviewees, Clark Abt of Abt Associates, put the question of why good evaluations are unused more succinctly. He maintains they are not used because:

- the prospective users don't know about them,
- if the evaluation is known, the user doesn't understand the results,
- if the results are understood, the user doesn't believe them,
- if results are understood and believeable, the prospective user doesn't know how to use them.

The process of use is a sequential one, and a gap at any given stage means nonuse. Abt's illustrations stem from his general experience as President of Abt Associates and specific cases. The latter include a recent conference on children's television run by the Educational Testing Service and attended by representatives of the commercial broadcasting industry. He maintains that many representatives were unaware of major research in children's television. Some of those who are familiar distrust it, partly because they distrust the social scientist. Irrespective of trust they argue that they cannot use it but present no evidence. This is despite remarkable experiences of "Sesame Street," Mr. Roger's Neighborhood," "Electric Company," and other programs which have been evaluated well. Abt maintains that cowardice accounts for some nonuse. For instance, the Follow-Through evaluations suggest that poor program variations should not receive support. They continue to receive support, however.

Abt's observations are not inconsistent with our findings.

Timeliness

The complaint that information, including evaluations, is not timely occurred frequently at the federal level, less often at state and local levels. The charge has not been confined to USOE's Office of Evaluation and Dissemination, of course. It has been leveled against the U.S. General Accounting Office, judging by Frederick Mosher's recent history of the GAO and our conversations with Congressional staffers. The charge is not especially well documented. Individuals may say that information is not timely but when asked for detail admit that they base the statement on one or two major examples, rather than an average. The absence of documentation does not obviate the point: late reports are damaging.

We believe there are several potential causes of the delays. The first is clearance of reports for release by the Office of the Secretary discussed in Chapter 5. The process has been time consuming, e.g. clearance of a recent report on utilization of evaluations took nearly five months. Delays of two to three months have not been uncommon. The second is clearance of questionnaires and the like at the start of the study. The process can involve six months before approval to begin information collection. Such delays can be devastating to short-term studies and formidable to long-term ones. New rules for both clearance processes have been created since 1978, and it is too early to tell whether they will have much effect. A third cause of delay is the sheer difficulty of carrying out a complex study, in which all obstacles cannot be anticipated, according to schedule. We are aware of no formal attempt to identify reasonable ways to accommodate the problem. Finally, reports are often complicated, require time to digest, to transform them into language which can be easily understood, and to identify ways in which results can be used.

At the national level, the charge that information is irrelevant has been leveled against evaluations by USOE, NIE, and the GAO. The charge is rarely made explicit. It generally concerns whether questions someone wanted to see answered were answered. It is a legitimate complaint in the sense that some evaluative questions are virtually unanswerable in particular cases, the agency is unwilling or unable to state this plainly, but the project is executed nonetheless. The complaint is gratuitous in that some Congressional staffers seem not to be willing or able to make a judgment about relevance until the final report is submitted. That judgment comes too late of course to do any good but does prevent embarrassment, often amicably.

At least part of the argument about relevance stems from failure to identify the pertinent audience in evaluation. Everyone's interest cannot be served. And if one group does capitalize on results and another doesn't broad claims about general utility are a bit silly. To understand what appears to be sensible practice, consider that the NIE Compensatory Education study focused on questions Congress wanted addressed, and there was sufficient negotiation to assure that questions remained relevant. The same questions were considerably less relevant to managers than to Congress, and the answers do not appear to have been used much by federal, state, or local managers.

Credibility

The matter of credibility emerged during our interviews. One aspect of the topic concerns the view that evaluations conducted by staff responsible for the program are simply not as trustworthy as an evaluation conducted by an independent agent. A second aspect concerns the view that criticism of an evaluation automatically implies that the evaluation is not credible. Neither view is particularly new, of course. The issue emerged in evaluation of medical regimens, e.g., coronary bypass operations. And the alternative strategies for reducing criticism are fairly well understood by researchers if not by users of the research: independent, competent critique, replication, and recognizing the simple fact that no single study is ever perfect and it is the accumulation of knowledge that is important. Trust in the evaluator is clearly important where the audience is unable to understand the evaluation, and lacks the resources for independent review. Neither trust nor intelligent mistrust can be built instantly at any level of government. And this is one reason why stability of evaluation units is critical.

Interest

Interest in an evaluation activity is a prerequisite for obtaining useful information. Without that interest, it is difficult to generate any information at all. The incentives for providing support are related to incentives for using the information: if the individual or institute can benefit, then the likelihood of the information being collected and used increases.

We are aware of few formal field tests of alternative systems for encouraging the use of the results of high quality evaluations. There has been sufficient experience to make some judgments about generating interest. It is sensible to expect audiences to be more likely to use information when they have had an active role in planning the evaluation and when they have made a commitment to using results, as in the NIE Compensatory Education Study and the OED Bilingual Education Evaluation. It is sensible to expect more interest when promotions and awards are based on use. But apart from experimental promotion systems we know of no systematic government personnel effort to pin rewards to the use of evaluation results. The less systematic efforts along these lines include explicit recognition of useful and useless evaluations in Congressional Reports and in Congressional decisions about agency budgets. They include systems such as the Joint Dissemination Review Panel and analogous systems at the state level which provide opportunities for funding based partly on quality and use of evaluation results.

Interpretability

Ability counts. Knowing about information, understanding it and how to use it are important.

Once the results of an evaluation are developed, the implications of the results must be educed. There are few formal procedures yet for doing so.

Moreover, contractors, agency staff, and Congressional staff appear to be mixed in their ability to do this. Indeed, some contractors have strong reservations about making recommendations based on their work. Simply because time available for evaluations is not sufficient to assay options and recommendations well. Regardless of ability, recommendations are always likely to be arguable to warrant several viewpoints, and if different proposals are made, they need to be synthesized.

Printed information is often insufficient for understanding. Moreover, there is no formal mechanism set up to provide oral reports on evaluations to members of Congressional staff to assure understanding. Nor is there any formal mechanism to present the results of independent appraisal of the evaluation itself. Agency staffers are frustrated by their inability to initiate contact with Congressional staff. Congressional staff are frustrated by paperwork burden.

Apart from the problems engendered by the difficulty of constructing recommendations, and difficulty of communicating them, the absence of regular meetings invites errors in interpretation. For example, Carl Wisler of OED notes that in 1979 a subcommittee staffer on the Senate appropriations committee misinterpreted a finding from the Development Associates' 1978 report on bilingual materials, and recommended cuts in funding for materials development. This misunderstanding was later clarified, and funding was restored.

Lois-ellin Datta has noted that, even among agency staff, mis-statements and overgeneralizations creep into executive summaries unless there is sufficient review. For example, one state official mentioned to us that a visiting member of an OE team had expressed the opinion that the RAND study of federal programs supporting educational change had showed that innovative programs do not work. In fact, the study had shown some of the conditions under which such programs could be expected to work.

We do not know how serious this problem is. The matter was brought up by only three or four of our respondents out of some 70 at the federal level. But it was brought up by thoughtful people.

Finally, no formal mechanisms exists for routine following-up to determine whether recommendations are acted upon. Yet such information is useful for understanding how often recommendations have been considered, accepted, or rejected, for understanding conditions under which recommendations are most likely to be adopted, and ultimately, for understanding Congressional and agency performance.

Decision Options, Relevance, and Ability to Use Information.

In principle, one ought to be able to outline each possible major outcome of an evaluation and the decisions that one could make based on each outcome scenario. Specifying the range of possible decisions helps to anticipate, in an orderly way, how one could use the results of evaluation. It is a vehicle for tracking utilization after the evaluation is complete and decisions are made.

In practice, specifying the range of decisions is difficult or impossible because: (a) possible outcomes of evaluation are not specified in advance, or (b) the time is insufficient to specify decision options, or (c) the nature of decisions simply cannot be specified well before the information is collected, or (d) nobody is willing to specify options.

We have found few formal attempts to lay out decision options for any major evaluation, any federal or Congressional group. The exceptions include so-called evaluability assessments which do try to address the question of how the information will be used once it has been obtained. Exceptions also include evaluations that enjoy a long planning period. For instance, the NIE Compensatory Education Study staff took six months to clarify Congressional interests and there was at least some attempt to anticipate (but not specify on paper) how such information could be used. The process of specifying decisions within an agency is feasible at least at times judging by evaluability assessments which have been undertaken. Recent studies by GAO however suggest that it is all but impossible to complete evaluability assessments for broad aim programs because program goals are not clear. The consequence is that evaluation goals cannot be clear and neither can one specify possible outcomes and decisions clearly.

Regardless of whether decision options can be specified, there are often strong disagreements about what any given piece of reliable information implies. Consider, for example, a program found to have failed on most counts in meeting its objectives. At least one camp within the federal executive branch takes the position that more money ought to be put into the program then to make it succeed. The public wanted it: it should be made to succeed. A second camp will argue for its termination because it failed. Still a third camp will make the decision one way if the program is a demonstration project and the other if it was created as a service program. The point to recognize is that either decision is legitimate and that this does not imply the evaluation results are useless. The evaluation informs the decision, but the decision itself must be based on other values or theories of what one ought to do in the face of failure.

The business of specifying decisions is complicated as well because alternative failure and success are typically mixed. A program may foster reading ability, and also impede arithmetic ability. Children may learn no more, but parents may learn a lot. And so on. This complexity should not, we believe, impede attempts to specify outcomes or decision options. It is a persistent difficulty, and it's doubtful that we will make much headway by ignoring it.

Finally, innovative social programs do not succeed dramatically as often as we would like. Anticipating for small rather than large advances is prudent, and one can do so by exploiting high quality evaluation designs. Moreover, it is not unreasonable to prepare for failure of innovation by specifying what else will be tried if the current effort fails. Such contingency planning is a mundane exercise, but it is difficult to find more than honorific attention to the matter outside planned research and development efforts.

6.8 ILLUSTRATIVE CASE STUDIES

Simply asking an individual the question, "Are the results of evaluation used?" is a miserably inadequate device for understanding utilization. Flaws in recollection, the difficulty of verifying statements, fragmentary evidence, occasional deliberate deception, and similar problems argue for a case history approach rather than simple inquiry if the idea is to understand how evaluations are used. The problems of understanding use from self-reports are similar to those encountered in economic research on employment, medical research on use of health services, and the like. Consequently, we undertook detailed case studies of several evaluations.

We have two overlapping samples of studies: one which consists of studies completed by the Office of Evaluation and Dissemination in 1979 and 1978, and a second sample selected purposively. In the purposive sample, we asked federal agency staff members, Congressional staffers, and contractors to identify remarkable evaluations which had been used over the past three years. Then we tried to determine whether their contention about use was supported by evidence. The same strategy was used to identify interesting cases in the views of local and state staff. These studies then are illustrative, intended to complement the general description given earlier. The stress is on uses by the Congress, by management, and by oversight groups. Special efforts were made to secure documentary evidence. Most of the work of corroboration was done by telephone and mail. To facilitate independent verification of the evidence, we have included references to published documents and have acknowledged individuals who provided information.

The OED sample involved choosing those studies from the highlights of the OE Annual Evaluation Reports, for FY 1979 and 1978, that appeared from their description to involve evaluation. Studies of management *per se* were excluded, as were studies of economic projections.

Sample of Studies from Annual Evaluation Reports
1978 and 1979

Purposive Sample:

ASPE: Study of Impact Aid (in house)
Exploratory Evaluation of Follow Through (in house)
Fund for the Improvement of Postsecondary Education (NTS)

CE: Federal Program Supporting Educational Change (1977) (Rand)
Evaluation of Bilingual Education Projects under Title VII (AIR)
Evaluation of Follow Through Planned Variations (Abt)
Sustaining Effects Study (Systems Development Corporation)

NIE: Compensatory Education Studies (Paul Hill et al)
Title I testing (SRI)

ACYF: National Day Care Study (Abt)

Other: Uses by Congressional Budget Office
Uses by Parent Advisory Committees
Uses by Providence, RI School District

OED Sample

1979

Title I services to neglected and delinquent youth (Systems Development Corp.)

Study of magnet schools (Abt Associates)

Evaluation of Project Implementation Packages (American Institutes for Research)

Study of Campus-Based Aid and Basic Grant Programs (Applied Management Sci.)

Sex equity in vocational education (American Institutes for Research)

ESAA-TV survey of viewers (Applied Management Sciences)

Sustaining effects study (Systems Development Corporation)

1978

Evaluation of bilingual education projects under Title VII (AIR)

Evaluation of Follow Through Planned Variations (Abt)

ESAA aid to non-profit organizations (Rand Corporation)

Upward Bound (Research Triangle Institute)

Indian Education Part A (Communications Technology Corporation)

Exemplary programs in career education (AIR)

Case Study on the Use of Evaluation:
The NIE Compensatory Education Study

The National Institute of Education was mandated by Congress under Public Law 93-380, the Education Amendments of 1974, to examine "the fundamental purposes of [compensatory education] programs and the effectiveness of such programs in attaining such purposes." Dr. Paul Hill, director of the staff, worked with the administrative support of NIE which was required to report directly to Congress rather than through DHEW.

Congress wanted information on two broad subjects: the present operation of Title I at federal, state, and local levels, and the probable effects of changes in Title I legislation. The study staff selected 6 major topics, in consultation with Congressional staff, to be focused on in their report: delivery of Title I services, state and local administration of Title I, funds allocation methods, and effectiveness of various instructional techniques.

Evidence of the usefulness of the study is abundant. The House Committee on Education and Labor is explicit: "The Committee has found the quality of the research by NIE to be excellent and has consequently relied upon these reports in formulating amendments to Title I." (House report on HR 15, p. 5) And from the Senate Committee on Human Resources, we have: "The committee wishes to commend the National Institute of Education for the uniformly high quality of its study, as well as its timeliness, as it proved invaluable to the committee in the formulation of the Education Amendments of 1978." (Senate report on S. 1753, p. 7) The House Committee used the NIE report to organize the subject matter of the hearings on Title I.

The Senate and House reports cite findings as justification for the form and content of several Title I amendments. On occasion, the findings are so distinctive that their contribution to amendments admits of little doubt. The actions, statements, and proposed amendments that accompany citations of NIE findings are summarized in the following:

Effects of Title I on Recipients

Both the House and Senate reports explicitly acknowledge the study's positive findings, that services are delivered to appropriate children, that the program contributes to educational experiences of these children, and that Title I does enhance student achievement in districts in which the program is stable and well implemented. The House report asserts that: "All these findings can be contrasted with earlier studies which showed that disadvantaged students fall more and more behind in their achievement levels and become increasingly pessimistic about their ability to improve through education." (p. 7)

The Study then changed or reinforced some attitudes toward the effectiveness of compensatory education, a major impact of the information in and of itself. The House report used this information in its argument that Title I be reauthorized for another five years and that increased funding is warranted.**

Funds Allocation

Congress had altered the formula for allocation in 1974, such that only two thirds of AFDC children could be counted in a district's application for funds. Both the House and Senate reports cited the NIE finding that only 6.9% of participants qualified for Title I under the AFDC measure, as compared to the 60% figure Congress had been given in 1974. Moreover, AFDC recipients were unevenly distributed, with urban areas relying heavily on AFDC eligibles for Title I funding. These considerations led the committee to restore a full count of AFDC children to the formula for funds allocation. The proposed amendment became part of PL 95-561 (Title I, Part A, section 112).

The House report noted the Study's examination of alternative poverty formulas that would recognize needy individuals not currently eligible for Title I. Such alternatives would increase funding for the Northeastern and North Central states, while the share of the South would decrease slightly and the West would be unaffected. It was clear that the Committee had seriously considered the information presented, but they noted that no one index of poverty would perfectly capture the distribution of needy individuals. The Committee therefore proposed, among other amendments, that all funds up to the 1979 level of appropriations would be distributed according to the present index. These proposals subsequently became law (Title I, Part A, Section 111, paragraph 2 D).

The House report cited the Study's evidence that states differ in their criteria for allocating funds within counties to particular school districts. Because school districts are not coterminous with counties, the process of subcounty allocations can become very complex. The Study also found that several states pool grants for all counties and give funds to districts based on the states' own eligibility formulas. NIE concluded, "this practice violates the basic Title I statute and the regulations, but it could produce results wholly consistent with the intent of both." "Consequently" (p. 12), the House Committee proposed an amendment permitting States to allocate money directly to districts, ignoring county allocations. The Senate report proposed a similar amendment, and it became law (Title I, Part A, Section 111, (a) 2B).

The Study staff discovered that Title I funds per participating child were lowest in poor rural districts, because these districts have high concentrations of poor children, and because they are located primarily in states with the lowest education expenditures. Low levels of funding for general education worked against Title I efforts. This information in conjunction with other testimony was presented as the rationale for an amendment authorizing $400 million for supplemental grants to districts in counties with high concentrations of poor children. The Senate bill also contains this provision, and the Administration had also proposed it. It subsequently became law (Title I, Part A, Section 117).

The Study staff discovered that strong pressures exist within LEAs to increase the number of schools receiving Title I. The goal of concentrating aid on schools with the lowest income was not being met. NIE's finding that most demonstration districts chose to allocate funds to schools on the basis of student achievement criteria was noted. In expressing concern that funds might be diluted, the House report noted NIE's conclusion that demonstration districts had extended services to more students without reducing the intensity and quality of services, but they could not do so for long without increased funding.

Citing this and other evidence, the House and Senate bills proposed several actions. First, LEAs would be authorized to rank schools both according to poverty and educational deprivation. Poverty schools would be served in the order of their rank unless a school ranked lower in terms of the number of educationally deprived children. Second, regulations would be created to allow LEAs to serve schools ranked lower in terms of educational deprivation than schools ranked higher in terms of poverty. Third, LEA would be permitted to skip schools receiving special state or local services similar to Title I and like amounts. Finally, the bill required that once a measure of ranking is chosen by a district, it must be uniformly applied. These proposals became law (Title I, Part A, Section 122 and Section 123, (d)).

Services Delivered to Students

The Study found that less than 1% of secondary school students receive Title I services compared to 20% in elementary school. The House and Senate reports cited this finding and noted that many districts hesitate to adopt high school programs because they do not know the program types that would be legal. Both Senate and House Committees noted that the Commissioner should include in regulations legal models for issues that might arise due to use of Title I funds in elementary and secondary schools.

Both the Senate and House reports cite the Study's conclusions that, while Title I does not require or encourage particular instructional strategies, some state and local officials believe that HEW auditors prefer the "pull out" design. Both Senate and House Committees direct OE to develop regulations describing both "in class" and "pull put" designs for Title I administrators.

Both House and Senate reports cited the Study finding that a quarter of Title I students are assigned to homerooms exclusively for Title I students, and may be separated from higher achieving students for the entire school day. Both reports emphasize that Title I does not intend such segregation, which should be avoided. If it does occur, programs must show that Title I children are receiving their fair share of state and locally funded services.

The Study data revealed that parental involvement activities by 1976 had accounted for the largest expenditures for Title I auxiliary services, that a third of all districts surveyed had no functioning chairperson for a school advisory council, and that one fourth of districts had no council at all. Citing this and other information, the House bill revised the requirements for parent advisory councils. Two of these provisions became law. A third provision was altered in the final legislation (Title I, Part A, Section 125). NIE was also directed to conduct a study of the effects of parental involvement in Title I programs (Section 125).

The Study concluded that the "supplement, not supplant" provision for the distribution of state and local funds provided sufficient flexibility for districts. It also concluded that regulations promulgated by OE were not sufficiently comprehensive, because some of the tests for compliance were not set forth in the regulations and the regulations did not take the needs of administrators into account. Both the House and Senate reports cited these findings and required that tests of compliance on the supplanting issue be published in the Federal Register. Moreover, OE regulations were required to set forth legal models for non-supplanting of funds and to explain how these principles apply in day to day situations.

According to the House report, the Study concluded that an amendment was necessary to encourage states and school districts to give compensatory education students their fair share of services that are locally funded, establishing compensatory education services similar to those of Title I. An amendment for such comparability was introduced and later became law (Title I, Part A, sections 126 (c) and 131).

Both the Senate and House reports cite evidence from the Study that less than 4% of private school students in Title I districts are served by compensatory education services. The Study suggested that private schools may not be made aware of availability of Title I services or that public school officials may not design the services to take into account the needs of private school children. Both House and Senate Committees introduced two amendments to ameliorate this problem. First, equal expenditures for Title I children in private schools would be required, although the number of such children and their special needs are to be considered. Second the Committees proposed that the Commissioner be required to exercise his bypass authority to ensure prompt resolution of complaints of private schools dealing with Title I. In addition, the Committees urged strengthening of OE regulations to ensure that private schools become aware of the availability of services to their Title I eligible students. These proposals subsequently became law (Title I, Part A, section 130).

According to state administrators interviewed in the Study, most problems in coordinating Title I and state-funded compensatory education were due to a lack of clarity in interpretations and guidelines issued by Federal Title I administration, or to inconsistencies in Federal monitoring and enforcement. These findings, along with other testimony, led the House and Senate Committees to clarify language dealing with the requirements for obtaining an exemption from Title I's excess cost and comparability provisions. A new exemption would apply to state programs that are being phased in. Another proposal required that OE or SEAs determine in advance whether special state or local programs satisfy the requirements for exemption. These amendments subsequently became law (Title I, Part A, Section 131).

Both the Senate and House Committees concluded that when schools have a very high percentage of low income children, school-wide projects to serve them were in order. This conclusion was based in part on evidence from the Legal Standards Project, supported by the Study, that in such schools it is difficult to design "special" programs for this large majority which do not also serve the school as a whole. Both the House and Senate Committees therefore proposed an amendment permitting schools with high concentrations of children from low income families to use state, local and federal funds to design a single compensatory education program for all children. A revision of the proposal became law (Title I, Part A, section 133).

State Administration

The Study concluded that many states are not clear as to their authority in rulemaking, disseminating information, providing technical assistance and monitoring compliance. This confusion was attributed to ambiguity and difficulty of legislation and inconsistency of federal monitoring. Moreover, the legal framework for Title I has led some states to practices that are not in compliance. The House report clarified rulemaking language in proposals which subsequently became law (Title I, Part C, section 165). The Senate report discussed these findings under their proposed Title V, and the rulemaking amendment became law (Title V, Part A, Section 504). The House Committee also directed OE to clarify regulations and insure that states do not misinterpret their authority.

NIE found that states had implied authority to withhold funds, but that the manner in which they did so was "quite inconsistent." Citing this finding, the House report clarified language dealing with the manner in which states are to withhold funds. This proposal subsequently became law (Title I, Part C, section 186). The Senate report cited these findings under their proposed Title V, and the withholding amendment here became law also (Title V, Part A, Section 508).

The House report noted the NIE finding that Title I regulations dealing with state enforcement are scattered throughout the legal framework for Title I, thus impeding Congressional intent. The Committee urged the Commissioner to revise regulations for state enforcement such that they facilitated compliance and described options and the legal basis for sanctions. In addition, both the Senate and House bills included an amendment requiring states to submit monitoring and enforcement plans. This proposal subsequently became law (Title I, Part C, Section 171).

NIE concluded that most states require fiscal audits, as stated in their policies, but omit compliance audits. Moreover, audit resolution varies widely among states. Citing these findings, the House report introduced an amendment clarifying auditing and audit resolution responsibilities of state education agencies. This amendment subsequently became law (Title I Part C, section 170). The Senate report discussed these findings under their proposed Title V, and the auditing amendment becamse law (Title V, Part A, Section 509).

NIE suggested that states might become more effective in their responsibilities if the administrative set-aside portion of state funds were increased. The House report cited this and other testimony as justification for increasing the set-aside for state administration. The Senate report cited NIE's findings as justification for a different approach. The Senate bill consolidated state administration of Title I and Title IV under a new Title V of ESEA. The rationale for consolidation was that paperwork could be diminished and both programs could operate more efficiently. These proposals subsequently became law (Title V).

Administration

Both the Senate and House reports cited the Study's overall conclusion that Title I was generally well-administered at the federal level. They also noted problems that the Study had identified in the day-to-day administration of the program. According to the report, OE did not implement requirements clearly and consistently. SEAs and LEAs became confused about their own responsibilities as a result. OE particularly did not apply consistent standards in identifying violations of the "supplement, not supplant" requirements. The Study also found that the laws pertaining to Title I are not written clearly. Agencies must inquire of OE to get interpretations of the law. According to both the House and Senate reports, rewriting Title I legislation is an attempt to clarify the law.

NIE discovered that HEW had a poor record on audits. In light of this finding, both the Senate and House introduced amendments requiring the Inspector General of HEW to audit fiscal integrity and compliance of grantees and subgrantees regularly. NIE had also found that OE had acted in recent years on less than 5 percent of all monies involved in audit exceptions and that only 1 percent of this money had been recovered. The House and Senate bills outlined OE's responsibility to act on Title I audits and to recover money where necessary on a timely basis. Moreover, OE was directed to describe each step of audit resolution in regulations and to provide an annual report to Congress on audits. This last proposal subsequently became law (Title I, Part D, section 185).

The House and Senate reports cited NIE's finding that, although OE had developed a body of interpretation, guidelines and applications of regulations, this body of experience had not been assembled coherently together nor disseminated to state and local agencies. The House and Senate bills included an amendment directing OE to develop a policy manual for use by states and local agencies. In an appendix, the NIE report on administration includes recommendations for improving clarity of the federal legal framework. Some of these recommendations were included in the report and the Committees expressed support for them. The proposed amendment subsequently became law (Title I, Part D, section 187).

General Provisions

The House report noted that problems existed with current methods of demonstrating comparability between Title I services and those provided by state and local funds. It cited NIE's finding that the current method, using the non-Title I average, could lead to inequalities for Title I schools. The House therefore introduced an amendment allowing OE to waive this requirement in order to allow selected school districts to try alternative methods of demonstrating comparability. This became law (Title I, Part F, section 102). Testimony other than the NIE findings also was cited as justification in the House report.

The House report cited the finding that although procedures to set indirect cost rates were generally clear, there were two exceptional cases of a lack of clarity in regulations. The committee directed OE to clarify these regulations, which were leading to variations among states in setting rates in indirect costs for state administrative set-asides and for Title I programs for handicapped and neglected and delinquent children.

This listing is merely a content analysis of statements in the record of the two Committee reports. While it is not possible to establish a direct link between Committee action and the NIE Study findings in all cases, all the evidence points to the conclusion that the findings contributed greatly to the form of the bills. The NIE Study was highly useful in the sense that it contributed to at least 21 separate sections of the 1978 Amendments relating to Title I and Title V. Furthermore, the House committee directed the revision or addition of regulations to clarify procedures and policy in six other areas of Title I administration and funding.

In developing this case study, Paul Hill, former director of the Study, and Iris Rotberg, a co-principal investigator, were interviewed. Pertinent information was provided by Jack Jennings of the Congressional staff and Chris Cross, a former staff member.

In developing this case study, Iris Rotberg, deputy director of the study, was interviewed. Pertinent information was provided by Jack Jennings of the Congressional staff and Chris Cross, a former staff member.

References*

Hill, P. Evaluating education programs for federal policy makers: Lessons from the NIE Compensatory Education Study. In J. Pincus (Ed.) Educational evaluation in the public policy setting. Santa Monica: Rand Corporation, May 1980. (R-2502-RC)

National Institute of Education. Administration of compensatory education, Washington, D.C.: NIE, 1977.

National Institute of Education. Compensatory education servcies, Washington, D.C.: NIE, 1977.

National Institute of Education. Demonstration studies of funds allocation within school districts, Washington, D.C.: NIE,1977.

National Institute of Education. The effects of services on student development, Washington, D.C.: NIE, 1977.

National Institute of Education. Evaluating compensatory education: An interim report on the NIE Compensatory Education Study, Wasington, D.C.: NIE, 1976.

National Institute of Education. *Title I funds allocation: The current formula*. Washington, D.C.: NIE, 1977.

National Institute of Education. *Using achievement test scores to allocate Title I funds*, Washington, D.C.: NIE, 1977.

U.S. House of Representatives, Committee on Education and Labor. *Report: The Education Amendments of 1978, H.R. 15*. (95th Congress, Second Session, Report No. 95-1137), Washington, D.C.: U.S. Government Printing Office, May 11, 1978.

U.S. Senate, Committee on Human Resources. *Report: The Education Amendments of 1978, S. 1753*. (95th Congress, Second Session, Report No. 95-856), Washington, D.C.: U.S. Government Printing Office, May 15, 1978.

Footnotes

*Many of the NIE reports cited here are based on reports submitted by independent contractors. The contracts were administered by the NIE Compensatory Education Study staff and were directed toward answering questions that the Congress asked to be addressed in the Study. Contractors included Abt Associates (formula allocation), Educational Turnkey Systems, Inc. (administration), Lawyers Committee for Civil Rights Under Law (legal standards), Booze, Allen, Hamilton, Inc., Syracuse Research Corporation, Policy Research Corporation, and others.

**We are not aware of any intensive critique or secondary analysis of the Study's analysis of Title I program effects on children.

Case Study: Evaluation of Title VII Bilingual Programs

American Institutes for Research conducted an evaluation of the Title VII Bilingual Programs from 1975 to 1977. From 38 sites of bilingual programs in their fourth or fifth year of operation, classrooms were randomly selected from grades 2 through 6. Non-bilingual classrooms with comparable students were selected in or near the districts incorporating the bilingual classrooms. Children were tested for English comprehension and reading, mathematics, Spanish oral comprehension and reading, and attitudes. The evaluation was supported by OE's Office of Evaluation and Dissemination.

An interim report was submitted to Office of Education for the reauthorization hearings of ESEA in 1978. The findings included:

1) Hispanic students not in Title VII classes outperformed Hispanic students in such classes in English proficiency (with some variation across grade levels).

2) Hispanic students in Title VII outperformed non-Title VII students in mathematics. (In the final report, however, additional data and a new analysis showed that Title VII and non-Title VII students appear to do equally well in math.

3) Less than 1/3 of students in Title VII programs were of limited English speaking ability and generally did not exit the program when they became proficient.

4) Title VII Hispanic students had higher Spanish proficiency than did the non-Title VII comparison group, though the Title VII program may not have been the only contributing factor.

These findings appear to have been used by Congress to alter the Title VII program, judging from reports of the House Committee on Education and Labor, the Senate Committee on Labor and Human Resources, and other groups influencing the 1978 Education Amendments.

Congressional Use

The House report first cited the AIR findings, the controversy they produced, and the critique by NIE of the adequacy of the information. In spite of uncertainty over the findings, enough information existed from other sources to justify certain amendments.

First, the AIR study showed the inadequacy of the current definition of the target population. The definition, as of 1978, equated speaking ability with competency in English, omitting understanding, reading and writing. The AIR measures were English reading and understanding. The new House bill broadened this definition to include these concepts. The Senate report also contained this broadened definition, which became law (Title VII of PL 95-561, Section 703).

The House further specified the students to be served by stipulating the number of English-speaking children allowed in Title VII classrooms. The AIR finding that many bilingual classrooms contained a majority of English speakers was tempered by the consideration that English and non-English speakers might benefit from interaction in the classroom. The House therefore proposed that for "pull-out" bilingual programs, all students be non-English speaking, but for regular classrooms, that the number of English-speaking children be proportional to their numbers in the school. The Senate report, on the other hand, provided that a maximum of 40% of each Title VII classroom be English-speaking. Although the Senate applied the same rationale, the AIR report was not cited. The Senate provision became law (PL 95-561, Title VII, Section 703).

A change in law may have come about because of the AIR study, according to one of our interviewees, namely, the requirements that OE develop procedures for identifying children of limited English ability, and that OE develop models for bilingual education and models for the evaluation of these programs.

AIR found that 85% of project directors said that students did not leave the program when they became proficient in English. Citing this finding, the House report noted that the problem centered on assessment of children's ability to operate in the English-speaking system, and that assessment of this proficiency was left to the local system itself. The House report stated that arguments about whether bilingual programs should be "maintenance" or "transition" programs confused the issues. However, in the bill itself, there is a provision for "measurable goals for determining when those children no longer need assistance." Thus language was strengthened to indicate that English proficiency was the goal of the program. This became law (Section 721).

The House bill contained a general rule that federal assistance for bilingual education be limited to 5 years for any one project, with waivers for special circumstances. The rationale for this provision was that bilingual programs are expensive to start, but may not be so once they are established. The AIR report was cited, showing higher costs for Title VII bilingual programs, compared to non-bilingual programs serving Hispanic children. The five-year rule became law (Section 721).

The House report cited AIR evidence that, although 80 percent of teachers in their study had a regular or bilingual teaching credential, there was great variation in standards for being certified a bilingual instructor. The House required OE to conduct a study of the impact of teaching fellowships in bilingual education provided under Title VII. This amendment became law (Section 723).

The evaluation has also been used by Congressional support staff.

See the case study on the Congressional Budget Office.

Management and Executive Uses

After the Amendments of 1978 passed, OE contracted with the Southwest Regional Laboratory (SWRL) to identify entry and exit criteria for bilingual students. According to several of our interviewees, this contract was a direct result of the AIR findings on the number of English-proficient students retained in bilingual classes.

Secretary Califano was cited in Education Daily for January 23, 1979, as saying that "the administration will 'make it clear' to schools receiving bilingual money that the law mandates that children must 'learn English as rapidly as possible' and be taught other subjects in their native language only until they are proficient in English." Education Daily also quoted Ernest Boyer, OE Commissioner, as saying "the Education Amendments of 1978, PL 95-561, 'makes it explicit' that bilingual education is to be an English language-development program." Moreover, the Secretary directed OE to require that bilingual classes consist of 75% or more students of limited English ability.

State Use

The California State legislature used the AIR study. Assemblyman Dennis Mangers introduced a bill (AB 2400) that would have required that districts use multiple criteria to assess whether a student was ready to return to English classrooms. According to his legislative aide, the AIR study reinforced his feeling and other information that many pupils are not so limited in English that they could not benefit from English classes.

Judicial Use

The AIR study was also used in Cintron vs. the Brentwood School District in the State of New York. In this case, the Puerto Rican Legal Defense Fund sued the school district of Brentwood to gain more funds for bilingual education. The principal investigator of the AIR study was subpoenaed to describe the AIR findings. The court determined that the Brentwood district was operating a Spanish language maintenance program with no exit criteria for English proficient students. The court ruled that the school must have criteria whereby students judged proficient in English would leave the bilingual classroom. This case preceded the Education Ammendments of 1978. Although the Brentwood district was not in the AIR sample, information from the AIR study allowed the court to ascertain ways in which the bilingual program could be improved.

Publicity and Controversy

The AIR study generated a great deal of controversy and publicity. Even before the hearings on Title VII, several articles on the study had appeared in the press. This trend continues, with the study being cited on Prime Time Saturday, a CBS TV national broadcast, on May 12, 1980.

The study had many difficulties along the way. OE was taken to court after the contract was let, which used valuable time. The second year sample had to be drastically cut because of budget restraints. NIE critiqued the study, but other criticism is not nearly as conscientious and, at least at times, inept. The AIR principal investigator cited one example in which a critic had read, not the report, or even the summary, but a 4-page document produced for Congress.

The study is controversial in large part because of the evidence that federal programs were operating as native language maintenance programs. Although local districts are free to fund maintenance programs, the federal government will only fund programs that have as their primary goal the acquisition of English and eventual transfer of students to English language classrooms. This is a thorn in the side of some who regard the native language as being relegated to a secondary status. Moreover, transitional programs would soon lead to fewer bilingual classes in some cases. For these reasons the bilingual community attacked the findings vigorously.

In generating this case study we interviewed Malcolm Danoff, principal investigator for the AIR evaluation of Title VII bilingual education programs, Norma Partlow, legislative aide to Dennis Mangers of the California State Assembly, and Tetsuo Okada and Carl Wisler of OED. John Ellis, former Executive Deputy Commissioner of Education, was also interviewed.

References

Assistant Secretary for Planning and Evaluation. Report on evaluation utilization in the Department of Health, Education and Welfare. Washington, D.C.: Office of the Assistant Secretary for Planning and Evaluation, DHEW, April, 1980.

Berke, I. P. Utilization of educational program evaluation for federal policy making: The case of bilingual education. Paper presented at the Evaluation Research Society Convention, Minneapolis, October, 1979.

"Budget gives bilingual education $174 million for English emphasis," Education Daily, January 23, 1979.

Danoff, M. Evaluation of the impact of ESEA Title VII Spanish/English bilingual education program: Overview of study and findings. (AIR report 48300-3/78-FR VI), Palo Alto: American Institutes for Research, March 1978 (ERIC Report ED 154-634) 25 pp.

U.S. House of Representatives, Committee on Education and Labor. Report: The Education Amendments of 1978, H.R. 15.(95th Congress, Second Session, Report No. 95-1137), Washington, D.C.: U.S. Government Printing Office, May 11, 1978.

U.S. Senate, Committee on Human Resources. Report: The Education Amendments of 1978, S. 1753. (95th Congress, Second Session, Report No. 95-856), Washington, D.C.: U.S. Government Printing Office, May 15, 1978.

Wisler, C. E. Matching evaluation designs to information needs: The evolution of federal evaluation strategies. Presentation at AERA Convention, Boston, April, 1980. (Available from the author, Office of Evaluation, U.S. Department of Education, Washington, D.C..)

Case Study of the Use of Evaluation
The Fund for the Improvement of Postsecondary Education

In 1978-79, the NTS Research Corporation conducted an evaluation of the Fund for the Improvement of Postsecondary Education, a program then authorized under Section 404 of the General Educational Provisions Act. The evaluation was supported through a contract between the Office of the Assistant Secretary for Planning and Evaluation (ASPE), DHEW and the Corporation.

In testimony before the Subcommittee on Postsecondary Education, the director of the Study Sol Pelavin, concluded that the Fund was "extremely successful. . .when judged by any number of criteria." He stressed that such positive and relatively unqualified judgments about the worth of a program were relatively rare, making the Fund a remarkable exception.

The evidence that this evaluation was used stems from recognition of the evaluation in the Report of the House Committee on Education and Labor and from corroborative testimony of the contractor, the federal agency project monitor, and a Congressional staff member with responsibility for obtaining information about the Fund. The Committee Report, for instance, quotes Pelavin's testimony in justifying the "Transfer of the Fund from Section 404 of the general Education Provisions Act to Title X of the Higher Education Act to give the legislative visibility deemed appropriate by the Committee." (p. 56) The evaluation finding that the Fund was successful in meeting its objectives and Congressional intent along a number of dimensions also appears to have been used to justify authorizing an increased level of funding.

Information about the use of the evaluation was obtained from Keith Baker, a former staff member of ASPE in the Division of Education, Sol Pelavin of NTS Corporation, and Thomas Wolanin, Staff Director of the Subcommittee on Postsecondary Education, Committee on Education and Labor.

References

Pelavin, S. H. Statement on Fund for the Improvement of Postsecondary Education, before the Subcommittee on Postsecondary education, U.S. House of Representatives, July 13, 1979. (Based on research conducted under Contract #HEW-100-78-0128 for ASPE/DHEW). Durham, N.C.: NTS Research, 1979.

U.S. House of Representatives, Committee on Education and Labor, Education Amendments of 1980. (96th Congress 1st Session, Report No. 96-520), Washington, D.C.: U.S. Government Printing Office, 1979, pages 56 and 262.

Case Study on the Use of Evaluation: The Congressional Budget Office

We interviewed three members of the Congressional Budget Office whose primary responsibilities lay in educational budget policy. They each affirmed their use of evaluations generated by the federal government, but stressed that variations in quality of the evaluations preclude CBO's using results of some studies. The general uses cited in each instance involve development of an understanding of the degree to which program services are provided, whether the program has detectable effects, and costs. The index of use of evaluation employed in this case study is citation of evaluations in CBO reports to the Congress. We regard citation as a crude indicator of the use of evaluations in policy development.

So for instance, data from the Office of Evaluation and Dissemination's evaluation of the Basic Grants program was used in CBO's report to the Congress, Federal Student Assistance. The evaluation, conducted by Applied Management Sciences, provided CBO with preliminary information on how many needy students receive assistance from a variety of sources such as parents and federal grants. The data are important in understanding whether federal assistance can have much impact. The CBO report also relied on an Office of Evaluation and Dissemination evaluation of Upward Bound programs executed by the Research Triangle Institute. The specific use is in understanding that the program has some detectable effects, notably on influencing the likelihood that the needy student will pursue post-secondary education and no detectable effects on other behavior such as college grades. These OED evaluations are not the only reports cited. Descriptive data generated by the National Center for Educational Statistics, the Census Bureau, and other private and public agencies were also used. The report also cites individual university research efforts without specifying source of support. At least one of these, Alexander Astin of UCLA, has received OED support for longitudinal studies of college students' dropout rate.

The Congressional Budget Office's draft analysis of federal efforts for high schools exhibits similar dependence on evaluations. The documentary evidence is consistent with interviews. The early descriptive reports of the OED supported Sustaining Effects Study, for instance, is exploited to understand which states provide financial support for compensatory education beyond that provided through Title I and so understand where geographic needs may lie. The Congressionally mandated study of Title I, run by NIE, is used in a similar way. The early reports of the Sustaining Effects Study are also cited in CBO's description of the degree to which students who are needy are actually receiving appropriate Title I support and in describing what we don't know about support of high school students. No reference to any federally supported evaluation of vocational education in high schools is made. We believe the scarcity of reasonable quality studies accounts for this.

OED supported studies of bilingual education, especially an evaluation by American Institutes for Research, is cited with explicit reservations. Descriptive data on attitudes about federally supported bilingual programs are cited to assist understanding the conflict between federal and local views. The evidence on effects of the program is judged inconclusive by CBO staff. NIE supported efforts to estimate the effects of career education programs are mentioned but not cited explicitly in CBO's description of whether the program is effective. Similarly, OED supported evaluations of Upward Bound are cited in describing who is served and the effect of services but no explicit citation is provided.

CBO's report on day care services and the role of federal support relies heavily on seven major day-care studies. About 40% of the citations (pages 5-42) refer to recent Abt Associates evaluation supported by the Administration for Children, Youth, and Families. The remainder are most frequently syntheses supported by the Assistant Secretary for Planning and Evaluation. No OED or NIE studies are cited explicitly.

Apart from citation of this sort, CBO staff informed us that evaluations are cited in memoranda written in response to Congressional inquiry. We did not have the time available to examine the character and frequency of such use.

References

Abt Associates, Inc. National day care study: Preliminary findings and their implications. Cambridge, Mass.: Abt, 1978.

American Institutes for Research. Evaluation of the impact of ESEA Title VII, Spanish/English Education Program. Palo Alto, California: AIR, 1978.

Applied Management Sciences, Inc. Study of program management procedure in the basic grants programs.

Burkheimer, G. J., Riccobono, J. A., and Weisenbaker, J. M. Evaluation study of the Upward Bound program: A second follow-up. Chapel Hill, N.C.: Research Triangle Institute, November, 1979.

Congressional Budget Office. Childcare and preschool: Options for federal support.

Congressional Budget Office. The federal effort for high schools: A preliminary analysis. Staff Draft Analysis. Washington, D.C.: CBO, January 1980.

Congressional Budget Office. Federal student assistance: Issues and Options. Washington, D.C.: CBO, March 1980.

Decima Research. Study of the sustaining effects of compensatory education on basic skills. Report No. 2. Santa Monica: California, January 1978.

Case Study on Use of Evaluation:
The National Day Care Study

The National Day Care Study was executed during 1974-79 with the support of the Administration for Children, Youth, and Families at a cost of approximately $8 million. Conducted by Abt Associates, it focused on day care center operations and the services provided to children under five years of age.

Part of the study focused on the staff/child ratios in centers, on whether reducing such ratios would degrade quality of care, and on whether reducing ratios would decrease costs of federally supported day care. The questions have a direct bearing on day care standards and federal interagency regulations for child care. The National Day Care Study was one of four recent studies on the general topic and has been supplemented by independent informal interviews by federal agency staff. The uses to which results were put are remarkable in several respects. Some documentation on use is available and we outline the evidence below.

Federal Regulations

The NDCS had a direct bearing on revision of the Federal Interagency Day Care Requirements amended in 1975 to govern day care funded under Title XX of the Social Security Act.

In particular, proposals for new regulations were issued in the Federal Register in June 1979 (Volume 44, No. 117, pp. 34754-34781) and final regulations were issued March 19, 1980 (Volume 45, No. 55, pp. 17870-17885). The text of each makes it clear that the regulations are based heavily on the National Day Care Study.

> The study provided persuasive evidence that the 1968 FIDC requirements on size of the children's group and staff/child ratio in the classroom can be altered to reduce cost of care without appreciable loss of quality in care. The recommendations involve more stringent constraints on group size, the smaller being better generally, and relaxation of staff/ratio requirements, and three policy options on requirements were provided. The final regulations generally use age categories recommended by the NDCS to specify group and staff requirements, and employ one of three policy options for specifying group size and staff child ratios (pp. 17877-17878).
>
> The Study "found a critical relationship between specialized caregiver training and the quality of day care children receive." The proposed regulations incorporate the finding by requiring that all caregivers participate in "specialized training, on a continuing basis within six months of employment" (p. 34757). The same view is emphasized in the final regulations which require that "all caregivers without a national recognized child development credential regularly participate in training related to child care" (p. 17872). Further, the Study found no clear relation between any measure of welfare or development of children and prior educational attainment level and job related experience, and observed that it then "seems inappropriate to incorporate either into future federal purchasing regulations" (p. 161). The final regulation echoes this observation by not specifying entry level requirements for center caregivers.

The Study recommended that the federal requirements bearing on size of children's groups and staffing "be determined on the basis of scheduled enrollment of children rather than their actual attendance" (p. 34762). The recommendation was offered partly in recognition of confusion caused by children's absenteeism encountered at the Centers by Study staff. The final regulations permit designation of group size on the basis of either enrollment or attendance, the deviation from original NDCS recommendations being influenced at least partly by independent commentary on the proposed regulations arising from variations in state practices.

Oversight and the Congress

There was some use of the study by the Congress and the U. S. General Accounting Office in the following sense. Senators Robert Packwood and Russell Long asked the U.S. General Accounting Office for advice about day care regulations. In a written response of September 25, 1979, Comptroller General Staats recommended that the new regulations be based on the National Day Care Study and on the GAO's reanalysis of the work. Staats' letter reiterated NDCS study findings that early regulations on staff/child ratios and training were too stringent and good quality care can be obtained by relaxing these requirements within limits. Moreover, GAO's independent study of day care was used as a partial basis for verifying conclusions.

One rather important but difficult to verify use of the NDCS concerns resolving a ten year debate between Senators Long and Packwood and the Department of DHEW on day care regulations. The discussion concerned the stringency of regulations, the problem that some states could not meet requirements, and the Congress's continuing resolution permitting waiver of regulations for such states. The NDCS study provided a basis for resolving differences between the Congress and the agency, and for eliminating the need for the waiver resolution.

There is some indication of use at the general policy level in the following sense. The Council on Wage and Price Stability reviewed the study and complimented DHEW on its conduct in a Council document dated September 24, Docket No. 79-184-32. In its issues and options paper on federally supported child care, the Congressional Budget Office relied on seven major studies of day-care services including the NDCS. Their citation of NDCS accounts for over a third of the citations in their discussion of who is served, whether services are effective, and the need for services.

Verification of factual information on use of the National Day Care Study was obtained in telephone interviews with Richard Ruopp of Bank Street College, Jeffrey Travers of the National Academy of Sciences, and Herb Millstein of the U.S. General Accounting Office.

References

Coelen, C., Glantz, F., and Calore, D. Day care centers in the U.S.: A national profile 1976-77. Cambridge, Mass.: Abt, 1979.

Congressional Budget Office. Childcare and preschool: Options for Federal support. Washington, D.C.: CBO, September 1978.

Federal Register, volume 44, No. 117, Part III, June 15, 1979, pp. 34754-34781.

Federal Register, volume 45, No. 55, Part V, March 19, 1980, pp. 17870-17885.

Ruopp, R., Travers, J., Glants, F., Coelen, C., and others. Children at the center: Final report of the National Day Care Study (Summary Findings and their Implications). Cambridge, Mass.: Abt, 1979.

Staats, E. B. Letter to Senator Russell B. Long and Senator Robert Packwood. (FGMSD-79-48). Washington, D.C.: U.S. General Accounting Office, September 25, 1979.

Case Study on the Use of Evaluation:
The Joint Dissemination Review Panel

History

The present Joint Dissemination Review Panel (JDRP) was begun in 1972 by Dr. John Evans and Commissioner Sidney Marland as the Office of Education Dissemination Review Panel. The JDRP includes representatives of both the Office of Education (USOE) and the National Institute of Education (NIE). Its purpose is to provide an internal quality control mechanism over the use of federal funds for the dissemination of educational products and practices for which claims of effectiveness are made. This purpose is accomplished by requiring approval by the Panel that there is sufficient evaluative evidence to substantiate the claims made about the innovations. The programs and practices must also pass a pre-review screening for social fairness and absence of potentially harmful side effects which is done by the relevant Education Division program office. Most of the approved programs have been actively disseminated by USOE's Division of Educational Replication through the National Diffusion Network. Of 421 submissions to JDRP, 245 (57.7%) have been approved. Of those approved, 60% (149) were developed using Title I, Title III, or Title IV money. All approved programs are described in Educational Programs That Work.

Structure and Operation

The JDRP consists of 22 members, 11 each from USOE and NIE, who serve two year terms. Members are nominated by the Commissioner of Education and Director of NIE on the basis of their education and experience in evaluation and their practical knowledge of education. Panel members are selected on the basis of their interest and willingness to participate. There is no financial or "in kind" reward for serving. The Panel itself has no official status or budget within the Education Division although the Executive Secretary, Mr. Seymour Rubak, is a staff member of the USOE's Division of Educational Replication. Meetings of the Panel are scheduled by the Executive Secretary whenever 2 or 3 submittals have been received. Seven to nine Panel members are scheduled to attend each meeting. The meetings are very informal and open to the public. The vote on each submission is taken immediately after it is discussed and is also public.

Pre-review Screening

Individual program offices within the Educational Division are responsible for identifying and screening potential candidates. In addition to making preliminary judgements on the accuracy and adequacy of the evidence of effectiveness, the screening process is intended to insure that innovations are socially fair, free of race and sex bias for example, and present no potentially harmful side effects. The individual responsible for the screening signs a transmittal memorandum assuring the Panel that the screening was done and the innovation found acceptable. A USOE Assistant Commissioner or NIE Associate Director also reviews each submittal and approves it before it is reviewed by the JDRP.

Based on conversations with agency staff, this screening process seems to work well in general, although two exceptions were noted. The first relates to the quality of the evidence and its presentation in the actual submittal. Program offices vary in terms of their experience with the screening process and that affects the persuasiveness of the actual submittal, independent of the quality of the evidence. Federal agency staff members familiar with JDRP operations expressed the feeling that the quality of both the evidence and its presentation have increased over time. The second exception was that some products may have been approved which did not meet all of the criteria. For example, we were told of an instance in which a Texas school district complained that a JDRP approved program was not free of sex and race bias. Given the difficulty of applying these criteria, it is not unlikely that there will be some justified criticism of individual programs or products.

JDRP Review

The actual review process is based on the evaluation evidence collected to substantiate the claims of effectiveness made for the innovation. The evidence is presented to the Panel in the form of a ten-page submission which is supposed to include information about the innovation and its development, who funded it, what claims are made for the innovation and for what population, and what the costs of adoption would be. However, the major emphasis in the submission is to be on the evidence of effectiveness and factors affecting its credibility. The criterion used by the Panel are described in detail in the JDRP's Ideabook, along with numerous examples of both persuasive and unpersuasive evidence. Briefly the evidence must be reliable and valid. The effects must be both statistically and educationally significant. There must also be credible evidence that the observed outcomes were caused by the program and would not have occurred in its absence. Finally, the innovation should be generalizable to other sites at a reasonable cost.

It is clear from conversations with JDRP members and minutes of the JDRP meetings that establishing a credible argument for the causal link between the program and outcomes is one of the primary requirements for approval. The meetings provide Panel members with an opportunity to clarify any questions they have after reading the submittal and gives the individuals making the submission a chance to convince the Panel that their program is exemplary. The votes are almost always accompanied by comments which provide reasons for the Panel's decision.

JDRP Performance

Although no formal evaluation of the JDRP has been done, it is the feeling of the people to whom we talked that the Panel is serving an important function and serving it well. The standards it uses for reviewing evidence of effectiveness conform reasonably well to other related standards for judging evaluations, e.g., those issued by the Evaluation Research Society and by the Joint Committee on Standards for Educational Evaluation. What little evidence exists in published form (Tallmadge, 1977b; Crandall, 1975) suggests reasonable adherence to the criteria.

Evidence from an evaluation of the National Diffusion Network (Emrick, 1977) showed about 9000 adoptions of JDRP approved programs by local school districts. However, not much evidence of continued effectiveness is yet available. The cost of adoption has been about $4000 on average, while the average cost of program development was about $300,000. Major criticisms of the Panel involve three issues: the rigor with which the criteria are applied (c.f. Tallmadge, 1977a), the Panel's decision not to review proprietary products, and the issue of equity of opportunity in providing convincing evidence. Each of these issues is discussed by Datta (1977).

The information contained in this case study was obtained from the references cited below, minutes of JDRP meetings, and interviews with the following individuals:

Dr. John Evans, Assistant Commissioner, Office of Evaluation and Dissemination, USOE
Maryann Milsap, Senior Associate, Teaching, Assessment and Evaluation, NIE
Ann Bezdek, Office of Evaluation and Dissemination, USOE
Jeff Schiller, Assistant Director, Teaching, Assessment, and Evaluation, NIE
Seymour Rubak, Executive Secretary, JDRP, USOE

References

Crandall, D. V. Program validation: Four case studies. Andover, Mass.: The NETWORK, 1975.

Datta, L. The external implications of an internal review of effectiveness: The DHEW Education Division's JDRP. In The Education Division's JDRP: Three Papers. New York: EPIE Institute, 1977. (ERIC #ED156217)

Educational Programs That Work. San Francisco: Far West Laboratory, 1979. (ERIC #ED149441)

Emrick, J. Evaluation of the National Diffusion Network: Final Report. Menlo Park, Calif.: Stanford Research Institute, 1977. (ERIC #ED147327)

Evaluation Research Society. Draft evaluation standards, 1979.

Joint Committee on Standards for Educational Evaluation. Draft Standards for evaluations of educational programs, products, and materials. Kalamazoo, Mich.: Western Michigan University, 1978.

Tallmadge, G. K. Comments on Dr. Lois-ellin Datta's paper. In The Education Division's JDRP: Three papers. New York: EPIE Institute, 1977a. (ERIC #ED156217)

Tallmadge, G. K. Ideabook: JDRP. Washington, D.C.: U. S. Government Printing Office, 1977b. (#017-080-01824-1) (ERIC #ED148329)

Case Study on the Use of Evaluation:
The Impact Aid Program

An evaluation of the program for School Assistance for Federally Affected Areas (Impact Aid) was conducted by Lawrence L. Brown III, Alan Ginsburg, and Martha Jacobs, of the Office of the Assistant Secretary for Planning and Evaluation, DHEW. The study was initiated to assist the development of the Administration's proposal on Impact Aid for the Educational Amendments of 1978. The Chairmen of the House Appropriations Committee and that of Education and Labor had requested such a study in 1977, as had Senators Bellmon and Muskie. The study assessed revisions in the program embodied in the Educational Amendments of 1974, using data from FY 1976, the first year of implementation.

The ASPE researchers concluded first, that the government could justifiably eliminate Impact Aid to children associated with Federal activities that did not deprive districts of taxes. This would eliminate Impact Aid to districts with parents who worked in federal jobs out-of-district, and to districts with locally-owned public housing. However, the latter districts were often poor and urban. The ASPE group concluded that the revision of Impact Aid would only be equitable if Title I funds to such districts were increased.

A second conclusion was that the current formula for compensating districts permitted wealthy districts to inflate the amount of compensation, while a "floor" of half the national average expenditure per pupil did not reflect disparities of expenditures across states. Although 20% of funds were allocated to relatively wealthy districts that were only lightly impacted by the presence of Federal activities, heavily impacted districts showed a real burden from the loss of tax revenues, depended heavily on Impact Aid for funding, and were poor in comparison to the average for their respective states.

The ASPE researchers also found that Impact Aid might actually impede States from providing their own equalization money, because Impact Aid as currently formulated offered them no incentive to do so.

The ASPE report presented three comprehensive reform packages, which varied according to their departure from current practice and degree of cost savings. The Administration chose the second of these packages, which departed moderately from current practice. This set of recommendations included the following provisions:

1) That payments for children who work outside the county be eliminated.

2) That payments for children residing in low rent housing be gradually withdrawn.

3) That the method of computing compensation by local rates be eliminated for all but the most heavily impacted districts, and that the "floor" rate of 1/2 the national average be eliminated.

4) That a 3 percent "absorption" provision eliminate payments for federal children equal to 3 percent of the district's non-federal enrollment (eliminating lightly impacted districts).

5) That the "tier" system of funds allocation be eliminated.

Each of these provisions was offered by the ASPE report. Moreover, the Administration offered findings from the report as justification for the suggested provisions.

However, the Administration's proposal on Impact Aid was not adopted by Congress. The Senate Appropriations Committee did make some use of the information and expertise that the ASPE group had gained. According to Congressional staff member Sam Hunt, this information was useful in discussions about ways to cut the program. Although appropriations were finally cut, it is not possible to say that the ASPE evaluation had a great deal to do with it.

In creating this case study, we interviewed Lawrence Brown and Marti Jacobs, formerly of ASPE, and Sam Hunt.

References

Califano, J. Prepared testimony before the House Committee on Education and Labor. In Hearings on reauthorization of ESEA, Part 3, 1977.

Impact Aid two years later: An assessment of the program as modified by the 1974 Education Amendments.

Case Study on the Use of Evaluation:
Problems in Title I Testing

In a 1977 report, Sol Pelavin and Jane David of SRI International reanalyzed state and local school district data to develop recommendations about when to measure the achievement of children who participate in Title I supported programs. The options considered were (a) fall and spring testing (b) fall only (c) spring only. The origins of this work was earlier research by Thomas, Pelavin, and others showing that the expected achievement pattern of students changed drastically depending on whether summer was taken into account. The 1977 report acknowledges the National Institute of Education as the source of financial support for the work. The 1976 report, in which the problem of summer test score decline was originally recognized and described, acknowledges U.S. Office of Education, OED, support.

The 1977 report recommended that we base our estimates of student gains on the analysis of fall to fall testing data so as to recognize summer decline in test performance and so avoid inflating estimates of the effect of Title I programs. Their recommendation to school districts was that spring only tests be abandoned in favor of fall only testing, or fall and spring testing in the interest of encouraging uniformity of reporting and interpretability of test scores used to estimate the effect of Title I programs. The authors emphasize that summer decline in test scores may have some implications for summer schooling but do not make a formal recommendation about what to do in their final report.

Their recommendation that fall testing be made the regular basis for recognizing both effects of programs and summer declines (p. 59) eventually appeared in Section 124(g) of the Amendment to Title I of the Elementary and Secondary Act of 1965 (P.L. 95-561, The Educational Amendments of 1978). In particular, the law specifies that a local education agency's evaluation of students' basic skills be based on achievement testing "over at least a twelve month period in order to determine whether school year programs have sustained effects over the summer." Negative effects of the law have not been examined.

Part C of the same law directs attention to the problem of summer decline in test scores and to a mechanism which may help to control the decline. In particular, the law specifies that federal funds made available to states for improving local education practice can be used for summer bridge programs. The Pelavin-David report appears to have reiterated the problem of test score decline and to have informed Congress sufficiently about the character of the problem to produce an option for controlling it. The Pelavin and David report is not, however, the only one to recognize the issue. The origins of Part C lie partly in an earlier bill (H. R. 9968 submitted to the 95th Congress, 1st Session, 1977) by Representative Chisholm and that bill contained a similar provision for bridge programs based on the Pelavin-David work. We know of no serious secondary analysis of the study and its flaws.

References

David, J. L., and Pelavin, S. H. Research on the effectiveness of compensatory education programs: A reanalysis of data. Menlo Park, California,: SRI International, September, 1977.

Thomas, T. C., Pelavin, S. H., and others. Patterns in ESEA Title I Reading achievement. Menlo Park, California: SRI International, March, 1976.

Case Study on the Use of Evaluation: Federal Programs Supporting Educational Change

In light of the growing number of federal programs designed to produce changes in education through local grants and projects, the Office of Evaluation and Dissemination of OE awarded a contract to RAND Corporation in 1973 to examine the adoption of innovations in school districts. Paul Berman and Milbrey McLaughlin, the principal investigators, examined local adoptions of four "change agent" programs: ESEA Title IV-C (at that time, Title III), Bilingual Education, Vocational Education Exemplary Projects, and Right to Read. RAND surveyed school district personnel, from superintendents to teachers, in 293 districts, and studied 29 others in the field. The investigators interviewed state and federal officials involved in the change agent programs.

Berman and McLaughlin believe that although some school districts initiated projects to try to solve problems, others initiated projects primarily to obtain federal funds. Projects that were started in order to get federal money were not successfully implemented, because they did not have the commitment of local participants. Some of the problem-solving projects also fell apart because they did not choose a flexible strategy that accommodated the existing district organization. Four major factors influenced the continuation of the projects after federal money ended. These were: centrality of project goals to those of the district; demands placed on teachers to change (if light, more change); complexity of implementation; and consonance between the project philosophy and that of the district.

The findings of the RAND study influenced the reauthorization of ESEA Title IV-C in 1978. Citing the RAND study, both the House and Senate reports noted that

> An evaluation by the Rand Corporation of Federal funds for innovation found that these funds had a major effect in stimulating local districts to undertake projects that were generally consistent with categorical guidelines. However, factors at the local level resulted in successful implementation of only some of these projects and long-run continuation of even fewer. (House report, p. 61, Senate report, p. 50)

At a later point, the House report noted that a perceived need is for more innovative approaches to compensatory education. Former Secretary Califano had advocated allocation of funds for this purpose. Again citing Rand's conclusion that providing federal funds can catalyze local commitment to such projects, the House proposed an amendment such that half of additional appropriated funds in Title IV-C would be used for innovations in compensatory education. This proposal subsequently became law (Title IV-C section 431).

In noting the frequent lack of local commitment to continuing projects after termination of federal money, the House report notes that the period of federal funding should be made explicit, so that other districts could benefit from federal seed money. Citing the Rand report finding that few districts prepared themselves for termination of federal funding, the House report set forth an amendment specifying that funds for projects would be available for a maximum of five years. Beginning in the third year, the level of federal funding would start to decline. This amendment subsequently became law (Title IV-C, section 432).

Dr. Berman had testified before the House Subcommittee on Elementary and Secondary Education, presenting his evidence. In addition, several officials in HEW were instrumental in incorporating the RAND findings into the reauthorization decisions. Marc Tucker, currently an associate director of NIE, was involved in HEW policy analysis for the reauthorization. He reviewed many of the evaluations sponsored by OE to find some common threads for changes that might be made in the structure of programs. Of all the evaluations he reviewed, the RAND study was most valuable to give a sense of local services delivery. Although the legislative proposal he helped develop was not the one eventually submitted, the final proposal did retain ideas taken from the RAND report.

Brenda Turnbull was an assistant to Marshall Smith, then Associate Commissioner for Policy Studies, furing the period of the reauthorization hearings. She mentioned that her office was disturbed at the RAND finding that innovative projects were isolated from the rest of the school. They therefore inserted language in the Administration's proposal that local districts must make a commitment to spend some money on the project themselves, over time, while the federal share would decrease. Marshall Smith's office also inserted language in the proposal that projects should be integrated into the rest of the school.

Smith's office was also impressed by evidence from the RAND report and from an evaluation of compensatory reading programs, that the school building is the level at which improvement take place, more than the district or the classroom. They therefore inserted language in the Administration proposal that Title I projects as well as Title IV-C could involve innovations in compensatory education, implementing the building wide approach.

The RAND report has also been used in the Office of Education in the management of the National Diffusion Network. Specifically, the report gave managers of this network information on areas of assistance they could provide to local implementors-- the importance of adapting innovations to local circumstances.

The report has also been useful to states. The director of Title IV-C in the state of Kansas said that the RAND report had served as the basis for changes in the program. For example, Kansas did not allow funding of projects beyond 3 years, except for the dissemination of project ideas to other districts. The RAND report convinced the director that a fourth year of funding for the school district that developed the original project was in order, so that the innovation could become institutionalized in the district. In general, the Rand report confirmed the director's own information that little steps and gradual changes are more likely to be accepted by school districts. This is the advice that the director provides to local implementors.

Legislators in the state of California incorporated information from the RAND report in formulating two bills which were passed in 1977. The first, AB 65, required LEAs to adopt minimum standards of proficiency for students to receive degrees. However, these standards were to be developed locally, consistent with a state assessment framework. The rationale for local control, discussed by Assemblyman Gary Hart, was to be found in the RAND study: local development of projects leads to better implementation.

Bill AB 551 was also passed in California in 1977. It has become known as the School Improvement Act. It provided for staff development in LEAs. LEAs were to submit locally developed plans for development, and a waiver provision explicitly noted that requirements of the Bill should not stand in the way of needed and beneficial innovation. Again, Assemblyman Hart's rationale for the Bill referred to the importance of local development and explicitly to a paper derived from the RAND report by Milbury McLaughlin discussing successful and unsuccessful staff development programs.

One specific finding of the RAND report has received a great deal of publicity because it has fit the spirit of the times. The finding that local development of projects was superior to advice of outside experts was extensively cited by newspapers, for example. NIE's Lois-ellin Datta cites commentary and editorials on the RAND finding from Science News, The Daily Oklahoman, and the Rapid City (South Dakota) Journal.

However, Datta criticizes the specific finding that locally developed projects are more likely to be implemented, on the basis of the RAND data itself. A follow-up survey indicates that no experts, inside or out of the LEA, are particularly influential, but outside experts are perceived by teachers as being more useful among those projects that were successful in being adopted. Datta therefore believes that this particular conclusion of the RAND study should no longer be cited. However, this does not invalidate other findings of the study. No major secondary analysis has been undertaken.

The following people were interviewed in the development of this case study: Linda Bond, former legislative assistant to Assemblyman Gary Hart of California; Milbury McLaughlin, co-principal investigator of the RAND report on "Federal Programs Supporting Educational Change"; Phillip Thomas, Director of Title IV-C programs, Kansas State Department of Education; Mark Tucker, Assistant Director of NIE; Brenda Turnbull, former assistant to Marshall Smith, Executive Assistant to the Secretary of Education.

References

Berman, P. Prepared testimony before the House Subcommittee on Elementary and Secondary Education, for the Education Amendments of 1978, Hearings, Part 8, ESEA Consolidated Programs.

Berman, P., and McLaughlin, M. W. Federal Programs Supporting Educational Change. Santa Monica: RAND Corporation, 1977 (8 Volumes).

Datta, L. "Damn the experts and full speed ahead: An examination of the study of Federal Programs Supporting Educational Change as evidence against directed development and for local problem solving." Address presented at the Annual Meeting of the Evaluation Research Society, Minneapolis, Minnesota, 1979. (Available from the author, NIE.)

Hart, G. The California Pupil Proficiency Law as viewed by its author. Phi Delta Kappan, May, 1978.

McLaughlin, M. W. Pygmalion in the school district: Issues for staff development programs. Unpublished paper, RAND Corporation.

The Providence School Department, Providence Title I Evaluation Final Report, 1977-78, 1978-79.

State of California Assembly Bill No. 65, 1977.

State of California Assembly Bill No. 551, 1977.

U.S. House of Representatives, Committee on Education and Labor, Report: The Education Amendments of 1978, H.R. 15. (95th Congress, Second Session, Report No. 95-1137), Washington, D.C.: U.S. Government Printing Office, May 11, 1978.

U.S. Office of Education, Office of Evaluation and Dissemination. Impact of evaluation studies on OE Programs: National Diffusion Program.

U.S. Senate, Committee on Human Resources. Report: The Education Amendments of 1978, S. 1753. (95th Congress, Second Session, Report No. 95-856), Washington, D.C.: U.S. Government Printing Office, May 15, 1978.

Case Study of the Use of Evaluation: Follow Through

The Follow Through program was developed in part to sustain the effects of Headstart and other preschool programs on performance and other characteristics of children from low income families. Early appropriations were clearly not sufficient to meet a general objective of serving all relevant students. Consequently, the U.S. Office of Education shifted the Follow Through emphasis to assessing promising approaches to compensatory education and away from general service delivery which could not be met. Local education agencies involved in the program were then encouraged to select one of several models or approaches to educating participating children in the interest of better testing the effectiveness of the models. The models, developed by educational organizations and agencies and colleges, differed in emphasis, some stressing emotional and social development, others stressing direct instruction, still others the role of parents.

In 1977, Abt Associates produced a report evaluating the Follow Through Planned Variations Experiment. The principal conclusions were: There were greater variations between sites using a single model than among the models themselves. Follow Through and non-Follow Through children performed about the same, and in some cases Follow Through children's performance was higher or lower. Follow Through children were still scoring below grade level after several years in the program. Particular models, specialized approaches to education, were contrasted for their overall performance and there is some evidence that one such model had good results.

Judging from our interviews and documentation, the evaluation has had little effect on the Follow Through Program itself. The current director has not used the evaluation. Although Office of Education has attempted to reduce funds allocated to the program, this initiative has not been based heavily on the evaluation. Rather, OE's rationale is that the program was experimental in nature, and the experiment is completed. Although the program does face some budget cuts, it is not clear how these will be distributed among the Follow Through developers.

The evaluation has been discussed in public forums. Congressional hearings on Follow Through were conducted by the House Education and Labor Committee and by the Senate Human Resources Committee. A House staff member requested information about the findings from the study's project monitor at OED. The Senate heard primarily from the local directors of the Follow Through project models, and the committee's hearings then reflect only criticisms of the Abt Associates report (references below).

The evaluation was highly publicized. Several newspaper articles, cited below, emphasized a secondary, highly controversial finding that Follow Through models stressing basic skills performed, on the average, better than others, in spite of great variations among sites. This may have influenced, strengthened, or justified the "back to basics" movement in education, among readers of these articles.

A variety of factors appears to have impeded the clear policy use of the evaluation in any but minimal ways. First, no consensus was ever reached on the nature of the Follow Through program. Originally intended as a service program, its low level of funding led the Office of Education to regard it formally as an experimental program instead. Its original program design deteriorated in the face of management problems and pressure from program people, who still regarded Follow Through as a service delivery program.

Second, the evaluation has been subject to much debate from sponsors of the Follow Through models and from evaluation specialists. A critique, funded by the Ford Foundation, pointed to data analysis and conceptual problems. The critique confirmed some earlier criticism of the Follow Through evaluation by the General Accounting Office. Although sponsors of the various models were consulted about the objectives to be measured in the evaluation, criticism grew that the evaluation was not measuring the sponsors' objectives, some of which were so various and vague that any evaluator would be hard pressed to measure them. In addition, some staff members of Interdependent Learning Model and the Direct Instruction Model contended that the evaluators wanted to discredit the program. We have found no evidence that the evaluation was biased deliberately against the program.

Finally, Follow Through has had a vocal constituency pressing for continuance of the program. When cuts were threatened in the program, for instance, Follow Through parents and teachers wrote in to their Congressmen supporting the program. They clearly value the service delivery side of the program and that appears to be reflected in the legislative willingness to continue it.

In summary, the Follow Through evaluation provoked debate, partly for the right reasons and partly for wrong ones. Confusion about the nature of the program exacerbated problems about the role of the evaluation and acceptance of its findings. The evaluation has been exploited for rhetorical purposes, in the "basics are better" movement. It has also contributed to academic research, such as Cronbach's, in assessing the proper approaches to educational evaluations. And it has contributed to institutional understanding if we may judge by the GAO's description of lessons learned from flaws in the evaluation design and administrative procedures in evaluation. But we have discovered no clear linkage between results of the evaluation and major modification of the program apart from unsuccessful administrative efforts to reduce the program's funding level.

In obtaining details for this case study we interviewed John Evans, Director of OED; Walter Haney of the Huron Institute and author of a technical history of Follow Through's evaluation; Margaret Stivers, an administrator with the Kansas Follow Through model; Robert St. Pierre of Abt Associates and one of the investigators in the Abt evaluation; Rosemary Wilson, Director of Follow Through; and Carl Wisler of OED.

References

Anderson, R. B., St. Pierre, R. G., Proper, E. C., and Stebbins, L. B. Pardon us, but what was the question again? A response to the critique of the Follow Through evaluation. Harvard Educational Review, 1978, 48, 161-170.

Cronbach, L. J. Research on classrooms and schools: Formulation of questions, design and analysis (Stanford Evaluation Consortium, Occasional Paper). Unpublished paper, Stanford University, 1976.

Feinberg, L. "Basic" teaching methods more effective, study says. Washington Post, June 20, 1977, p. 1.

House, E. R., Glass, G. V., McLean, L. D., and Walker, D. F. No simple answer: Critique of the Follow Through evaluation. Harvard Educational Review, 1978, 48, 128-160.

Kilpatrick, J. J. Basics better in education, Cambridge group finds. Boston Evening Globe, July 1, 1977, p. 19.

La Tour, K. More high marks for the 3 r's: Basic skills found lagging in "open" classrooms. National Observer, July 4, 1977, p. 6.

Newsweek. Basic is better. July 4, 1977, p. 76.

Stebbins, L. B. (Ed.) Education as experimentation: A planned variation model. (4 volumes). Report by the U.S. Office of Education pursuant to Contract No. 300-75-0134. Cambridge, Mass: Abt Associates, Inc., 1977.

St. Pierre, R. G. The Follow Through debate. Curriculum Inquiry, 1979, 9 149-161.

U.S. General Accounting Office. Follow Through: Lessons learned from its evaluation and the need to improve its administration. (MWD-75-34). Washington, D.C.: U.S. General Accounting Office, October 7, 1975.

Weissman, C. Overview of the Abt Associates Inc. National Evaluation of Follow Through. Hearings before the Senate Subcommittee on Employment, Poverty, and Migratory Labor of the Senate Human Resources Committee, for the Economic Opportunity Amendments of 1978. May 14, 1978 (pp. 860-871).

Zimiles, H. Some notes on the evaluation of Follow Through. Hearings before the Senate Subcommittee on Employment, Poverty and Migratory Labor of the Senate Human Resources Committee, for the Economic Opportunity Amendments of 1978. May 14, 1978 (pp. 841-850).

Case Study:
Use of an Exploratory Evaluation of the Follow Through Program

The Office of the Assistant Secretary for Planning and Evaluation conducted an exploratory evaluation of the Follow Through Program in 1979. The evaluation found that there were conflicting views of the mission of the program that needed resolution. There were also disagreements about the management objectives of the program. Follow Through Central Office had no procedures for either producing effective services or for evaluating the effectiveness of those services. Finally, there was consensus in OE that Follow Through was unsatisfactory as it was functioning. Serious problems in the analysis were reiterated, however.

The Assistant Secretary for Education issued directives in 1979 that reorganized Follow Through at the national level and gave a new role to sponsors of projects, who were to concentrate more on local service and on knowledge-producing activities.

In addition, Follow Through was to develop performance indicators to monitor projects. A contract for this purpose has been awarded to Applied Management Systems. The development of these indicators was a major recommendation of the exploratory evaluation.

Another recommendation was that the research function should be returned to Follow Through itself, instead of OED, and should be staffed with qualified people. The Follow Through Office is waiting to implement this recommendation, which was authorized by Assistant Secretary Berry, but until organization of the new Department of Education is complete, this new staffing must wait.

The exploratory evaluation suggested three areas for further research: extending the Follow Through program through the 6th grade, varying levels of funding for Follow Through projects, and examining the effectiveness of self-sponsored projects. A contract has been let through the Follow Through Office to Boone-Young and Associates to carry out these studies.

To generate this case study we interviewed Rosemary Wilson, Director of the Follow Through Program, and supplemented this interview with existing written information.

References

Assistant Secretary for Planning and Evaluation, Final Report on Follow Through. Washington, D.C.: Department of Health, Education and Welfare, Office of the Assistant Secretary for Planning and Evaluation, Office of Evaluation and Technical Analysis, 1979, EDC 1012.

Assistant Secretary for Planning and Evaluation. Report on evaluation utilization in the Department of Health, Education and Welfare. Washington, D.C.: Office of the Assistant Secretary for Planning and Evaluation, DHEW, 1980.

Case Study on the Use of Evaluations:
The Providence, R.I. School District

Dr. Ron Visco has been the Director of Title I evaluation in Providence for close to three years. He was able to document a variety of instances in which the school district adopted recommendations he made on the basis of evaluations. Providence turns out an interim and a final evaluation report each year, documents the recommendations that were made, and the action that was taken on the recommendations. This also occurs for reports on the Title I program to the state.

Instruction

In 1977 and 1978, Visco investigated the effects of the cut-off criteria for Title I on reading and math and on services to children under Title I. There were two levels for selection criteria. If children scored between one half standard deviation and one standard deviation below the local mean on reading tests, they were given 60 minutes per week instruction time in reading under Title I. If they scored below one standard deviation under the local mean, they were given 150 minutes of instruction per week. For math, these criteria were similar, except that if a child were receiving 150 minutes in reading, he or she could receive no more than 60 minutes in math.

Dr. Visco found that for both reading and math, 60 minutes of instructional time had, at best, no effect, and possibly even detrimental effects. On the other hand, 150 minutes appeared to have a positive effect although there were some problems in interpretation. Dr. Visco therefore recommended that all children would receive 150 minutes of instruction, and that in math, the 60 minute service category not be considered again. The Providence Title I program has instituted this change.

In middle schools, the reading instruction involved two 45-minute periods per week. The results were poor. Visco recommended in 1978-79 that the middle schools increase instructional time. This recommendation was based on interviews and on data. Providence adopted this recommendation in 1979-80, increasing the number of periods to 3.

In 1978-79, a major debate concerned whether Title I classes should pull children out of their regular class to a resource room, or whether the Title I teacher should use the regular classroom, but teach Title I students in a corner. In 1978-79, a decision was made by Title I administration that almost all Title I instruction would occur in the regular classroom. This created problems. First, almost all schools have excellent facilities in their resource rooms, but these materials are not brought by teachers into the regular classroom, for a variety of reasons. Second, conducting two classes in one room can be extremely noisy, or uncomfortable for the Title I students. Visco administered questionnaires to Title I and regular teachers, and interviewed the Title I teachers. At the end of the questionnaire, he asked what would the teacher suggest about where Title I instruction should be performed. Of 131 questionnaires, only 5% responded that the regular classroom was preferable. 51% believed that the resource room was preferable. Title I teachers overwhelmingly preferred the resource room, or at least having a choice. Principals had more mixed opinions. Visco talked to 15 Title I students and each preferred the resource room.

Some Title I staff opposed distribution of the questionnaire, teachers filed grievances because of the rule, and in general the issue was very hot. Visco recommended several things: that more flexibility be allowed; that the location of instruction be based on the physical structure of the school, size of classrooms, size of regular and Title I classes, etc. Moreover, under some conditions, he recommended that the Title I teacher be allowed to remove children from the regular classroom at his or her discretion. The issue is still not resolved completely. However, Title I teachers are given more flexibility.

Visco conducted interviews of Title I teachers and determined that they were frequently assigned additional students, which could potentially impair learning. He recommended that Title I assign no more than 7 students per teacher, and that instead of being forced to take more students, teachers agree to having more students. Providence has adopted this proposal.

Visco also learned that Title I teachers wanted more workshops on reading and math. He recommended that such workshops occur regularly and no fewer than 4 per year. Providence adopted this recommendation.

His recommendation this year on the central high school were almost all adopted. For example, he recommended guidelines on the number of unexcused absences from Title I classes. This had been administratively difficult for the liaison person. Other recommendations included entry and exit criteria for the reading program and better selection criteria. He recommended an increased number of male tutors and involving the reading instructors in the selection of tutors.

Frequently, his recommendations involve modifying the program objectives which do not make sense or may be unrealistic. For example, Title I's objectives could only be reached if they performed evaluations with all three models. They therefore changed their objectives when Visco pointed this out to them. The Early Childhood Program wanted 80% of their children to pass each of the ten subtests for a particular test. Visco pointed out that this was not sensible in light of national norms that indicated that only one-third of non-Title I children in the norming group were able to pass any single subtest.

Classification and Achievement Testing

The evaluator discovered that clerical errors were preventing eligible children from receiving Title I services. He recommended that a computer search for children based on their scores be instituted, rather than relying on clerks to look for such children. The Title I program adopted this recommendation by having the testing company provide them with rankings of the children by their scores, as well as alphabetically. When Visco examined the error rate in 1978, after this procedure was instituted, he found no clerical errors in math, and only 4 in reading, out of a sample of 450. Only 1 of these errors prevented a child from receiving Title I services and this was remedied. In 1979, there were no errors at all.

Providence's Title I program has a challenge system, whereby a teacher can challenge a student's scores on the CTBS, the test they use for the criteria for Title I, if the teacher believes that the child is actually performing worse than his score would indicate. Visco examined major problems in misclassification of children in grades 1 and kindergarten. The reason for the misclassification was that the test was given during the first week of school, when it would be least indicative of the student's achievements. The test was also administered improperly, according to Visco, quite frequently. Visco therefore recommended that the challenge practice be extended to kindergarten and first grade, where they had not been previously. He also recommended that the challenge period be extended, in general, because he felt that it was too brief and also occurred too early in the school year for many teachers to be able to respond. Providence adopted this recommendation. In 1978-1979, they extended the deadline of the challenge period. In 1979-1980, challenges were permissible all year.

Visco felt that one of the biggest problems was the selection instrument for kindergarten children, called SEARCH. Norms for the test had been based on children at the end of kindergarten and beginning of first grade. However, Providence used the instrument at the beginning of kindergarten. This was inappropriate for several reasons, e.g., of the children receiving Title I, 98.6% got a score of 0 on one subtest. Moreover, the cutoff scores in the test had norms for children of, for example, 70 months, but these were children at the end of kindergarten, not the beginning. Visco recommended that the school district eliminate SEARCH, or failing that, get rid of the current cut-off scores in the test, and obtain local norms for determining the cut-off scores. Providence did this.

In several grades, the Title 1 cut-off scores were below the level that one would obtain by chance, merely guessing the answers. Part of the problem was that cut-off scores were developed in the district as a whole, but were used in Title I schools only. Visco therefore offered options to solve these problems. However, this year, almost all the Providence schools now receive some Title I funds, making the contribution of non-Title I schools trivial. In this case, his recommendation was irrelevant, although the guessing problem remains.

One of Providence's programs is English as a Second Language. The test they used to determine children's eligibility for the program was the Tests of Proficiency in English. This test was inappropriate, because it was developed in Wales and normalized on East Asian children. The language tapes relied on people with British accents and the picture materials referred to British objects and customs that Providence Title I children could not understand. His recommendation was to discontinue this test, and Providence complied.

Some of his recommendations in 1978-79 became irrelevant. For example, he made some recommendations dealing with the Metropolitan Readiness Test for first graders, but as the Title I program for this grade was cancelled, the recommendation was no longer relevant.

In 1977-1978, almost all the evaluator's recommendations were accepted and implemented by Providence. In 1978-1979, fewer were, but the number of recommendations he put forward was greater. In 1978-1979 he put forward eight recommendations on student selection, five on challenges, ten on process and instruction, seven on English as a Second Language, two on the school clinic, and sixteen on the Central High School program. This may have been too much for the administration to implement, or touched on too many hot issues. Interestingly, in 1977-1978, there was no state requirement to report the evaluator's recommendations and the Title I response. In 1978-79, there was such a requirement, yet in 1977-78, he feels that more recommendations were implemented by Title I.

Visco mentioned several factors that may have influenced his success in getting his recommendations implemented. One is the possibility that the state requirements about reporting recommendations and Title I responses may have moved Title I to adopt more. A second factor may have been that Visco is emphatic and persistent to get changes. Also, he has learned to speak to the Administrators informally first, before they see the recommendations in writing -- people like to be consulted. He regards merely "suggesting" changes as insufficient -- they won't pay attention unless a change is a formal recommendation.

References

The Providence School District, Evaluation Department. Providence Title I evaluation final report/1977-1978.

The Providence School District, Evaluation Department. Providence Title I evaluation final repot/1978-79.

Case Study on the Use of Evaluations
Parent Advisory Committees

Title I legislation provides that Parent Advisory Committees (PACs) shall be given "responsibility for advising [the LEA] in planning for, and implementation and evaluation of, its programs and projects" under Title I. However, the law does not specify how parents are to advise local education agencies. Our site visits and independent work confirm that there is a wide variation in the extent of parent involvement with and use of evaluations. In some districts, the Title I staff or evaluator merely provides parents with reports. In other districts, PACs are more actively involved in the evaluation itself, reports are interpreted to them, and so on.

We interviewed several people with national advisory roles bearing on PACs. These interviews gave us a perspective on the variation in parental involvement. For example, PAC's in South Carolina, Georgia, Alabama, and Mississippi have not used evaluations, to the best knowledge of Hayes Mizell, director of a group offering technical assistance to PACs. However, access to reports is a major problem in these states. Moreover, the evaluations frequently contain technical jargon that obfuscates the content. School administrators challenge negative findings about programs on the basis of the methodology or measurement used. Parents do not have the training to judge for themselves the quality of the evaluation.

On the other hand, the Washington, D.C., district-level PACs are extensively involved in evaluation, according to Tom Heatley. His organization has provided technical assistance to PAC's in 40 Washington schools.

Finally, there are examples of extensive use of evaluations by PACs. The Providence, Rhode Island, PAC use evaluation frequently, according to Constance Gomes, former chairperson of the PAC. In 1979, for example, the PAC noted the finding from evaluation that reading scores in middle school were poor compared to those in elementary school. The short time spent on reading in middle school was one probable cause. The PAC pushed for adoption of the recommendation that time in reading be increased. The LEA adopted the recommendation.

In the same year, the Providence PAC strongly recommended that children remain in the program on English as a Second Language for longer periods of time. This position was based on the evaluation finding that children were not reaching their objectives by the time they exited the program, but that children who remained in the program longer were more successful.

In 1973 evaluation revealed that the reading program was ineffective in Providence. The PAC advocated a new model for the program. The LEA instituted a new model.

More generally, the Providence PAC is involved in selecting the evaluator as well as objectives for evaluation. The PAC reviews both the interim and final reports, and advises the State Department of Education as to whether the evaluator's recommendations have been implemented. The Providence district code includes the provision that PACs must be involved in making recommendations on the basis of evaluation. Rhode Island in general does not have such a provision -- the PAC in Providence itself instituted this provision.

When Ms. Gomes was chairperson, the PAC and the evaluator would sometimes form teams of on-site visitors to interview parents about the progress of their children. Some parents have been members of the PAC for as much as 10 years, and have developed some sophistication about evaluation, as compared to most PAC members elsewhere.

Relations with the LEA have not always been harmonious in Providence. On one occasion the PAC took the district to court to gain access to evaluations. Now such access is a matter of course. On other occasions, the PAC went public with negative evaluations and embarrassed the LEA.

Other active PACs, according to Ms. Gomes, are the Wilmington, Delaware,and Springfield, Massachusetts,PACs. California has many active PACs.

Our respondents identified several major barriers to the use of evaluations by PACs. Dr. Robert Koff mentioned the natural resistance of administrators to involving parents in decisions that they regard as theirs. He also mentioned that, relative to parents involved in other educational programs, such as migrant education, it is frequently difficult to get Title I parents actively involved. Both Dr. Koff and Constance Gomes mentioned that teachers' unions have some antipathy to parental involvement in decisions. Unions are opposed to the comparison of classrooms or schools or districts on performance. It has therefore become difficult to make public such information, which according to Gomes was a major tool in increasing the accountability of the district.

Our respondents made several recommendations for increasing parental use of evaluations:

1) More training should be provided to parents in understanding the basics of evaluation. This is feasible to do and it is not necessary to involve the abstruse technical aspects. For example, Tom Heatley's organization gives parents a basic rationale for each of the three RMC models. TACs could serve this function. However, it is not currently part of their mandate, and they have not been very useful to date, according to Ms. Gomes and Mr. Mizell.

2) Evaluations should be worded in plain language that parents can understand. Their need for information involves knowing the objectives, how the objectives were met, and if they were not met, ways of trying to meet them in future.

3) Access to reports must be improved. Although the law now stipulates such access, there are still problems in some LEAs.

4) Involve the parents in evaluation from the beginning. If they participate in the choice of evaluation objectives, their questions are more likely to be answered. Such participation involves reading the Title I applications, understanding objectives and methods to reach educational objectives, and the basics of evaluation.

In developing this case study we interviewed Constance Gomes, member, National Advisory Council on the Education of Disadvantaged Children and Education Specialist, Rhode Island Legal Services; Tom Heatley, Executive Director, National Coalition of ESEA Title I parents; Dr. Robert Koff, member, National Advisory Council on the Education of Disadvantaged Children, and Dean, School of Education, SUNY Albany; and Hayes Mizell, Chairman, National Advisory Council on the Education of Disadvantaged Children, and Director, Southeast Education Program of the American Friends Service Committee.

Case Study on the Use of Evaluation: The National Study of Vocational Education Systems and Facilities

In a 1978 report, a Westat, Inc. study under the direction of Allen Woodruff provided the first comprehensive information on vocational education systems. The organization, governance, and financing of state systems were studied. General characteristics of 6,660 institutions were described, including geographical distribution, facilities construction patterns, numbers and distribution of shops and laboratories in each institution, condition of facilities, and accessibility of vocational education to the handicapped. According to Woodruff, the study was a pioneer effort that revealed many gaps in knowledge. It provided a data base for the uses listed here.

The report has been useful to interest groups, in that it has been cited in Congressional testimony by Gene Bottoms, President of the National Vocational Association. He cited information from the report in Oversight Hearings on Vocational Education, April, 1979; in hearings on the Youth Employment Act of 1979, on June 26, 1979; and in hearings before the House Appropriations Committee, April 26, 1979.

It has been useful in meeting legislative demands for information in the following sense. Several Congressionally mandated studies have made use of this report or of the data base derived from it. Charles Benson has made use of the Westat sampling strategy in examining educational financing of Vocational Education. Laurie Harrison, Director of the American Institutes for Research study of sex equity in Vocational Education, has made use of Westat's sample also. The National Center for Educational Statistics was able to delete from its Vocational Education Data System (VEDS) the component involving data on facilities in Vocational Education, because Westat had already collected the information.

The report appears to have been useful in federal management as well. Woodruff was also asked to assist the Department of Labor in their efforts on behalf of the Task Force on Youth Unemployment. He provided them with data in Volume 1 of the report, dealing with the extent to which the need for Vocational Education in urban areas was not being met. Woodruff and Dan Dunham, Assistant Commissioner of the Bureau of Occupational and Adult Education, analyzed unpublished data from the survey to address questions about the distribution of facilities, real levels of utilization, and the capacity of urban schools to handle an additional load in vocational education. Dunham has also used information from the Westat report in Congressional testimony on appropriations for the Bureau of Occupational and Adult Education in March of 1979 and 1980. (He has also used the information in speeches, especially the information on access to vocational education facilities).

Dr. Leroy Cornelson, formerly Director of Compliance and Grants in the Bureau of Occupational and Adult Education, has used information from the Westat report in several ways. He developed the 5 year and 1 year budgets for Vocational Education for the administration. He also used information in the report relating to the governance of the administration of Vocational Education in the states. For example, the report allowed Cornelson to assess the effectiveness of the provision that a "sole state agency" retain control over policy and decision making in Vocational Education. He used this information in outlining issues for the reauthorization of Vocational Education legislation.

The national study appears to have been of some interest at the state and local level. Final reports from Westat were purchased by 38 state agencies and 30-50 state advisory councils in Vocational Education. States have used the data as a resource in changing their laws and policy. For example, Georgia sent a representative from the joint legislative committee to discuss the implications of the research with Woodruff. Kentucky did the same. Woodruff has served as a consultant to a blue ribbon panel studying the Vocational Education system of the state of Iowa. Maryland contracted with him as a consultant on educational finance, to compare their resources in Vocational Education to those of other states. While these state uses are not documented, Woodruff is fairly sure that their thinking about their Vocational Education systems changed as a result of considering the report.

In developing this case study we interviewed: Dr. Charles Benson, of the School of Education at the University of California at Berkeley; Dr. Robert Calvert, Branch Chief, Adult and Vocational Surveys and Studies, NCES; Dr. Leroy Cornelson, Director of State Programs, Bureau of Occupational and Adult Education; Dr. Dan Dunham, Assistant Commissioner, Bureau of Occupational and Adult Education; Dr. Laurie Harrison, Director of the AIR study of sex equity in Vocational Education; Dr. Edward Rattner, Project Monitor of the Westat study; and Dr. Allen Woodruff, Principal Investigator for the Westat study of Vocational Education facilities.

References

Oversight Hearings on Vocational Education before the House Committee on Education and Labor, 1979.

Hearings before the House Appropriations Committee, April 26, 1979.

Westat, Inc. National Study of Vocational Education Systems and Facilities (3 Volumes). Rockville, MD: Westat, Inc., 1978.

The Youth Employment Act. Hearings before the House Committee on Education and Labor, 1979.

Case Study: Use of an Evaluation of Services to Neglected and Delinquent Youth under Title I

Systems Development Corporation was given a contract to examine compensatory education programs for children in state institutions for neglected and delinquent youth. As of this writing, their final report is being completed. An early volume of the report contained descriptive information about services and the children being served. An important finding was that only one half of Title I eligible students were being served. Moreover, Title I students received less instruction than did non-Title I students in institutions.

The descriptive information was used in the Education Amendments of 1978. The House Report cited findings from the study to demonstrate concern over problems with the program. After citing the high illiteracy level among such students, the report noted that "A national evaluation of the program found that about 56 percent of participating students felt the Title I classes in reading and math were teaching them more than other classes they had taken in these subjects." (p. 39)

To insure that Title I students would receive more instructional time, an amendment was created to emphasize that in institutions, Title I funds are to supplement, not supplant money for education provided by the states. According to one former Congressional staff member, the findings of the report influenced this decision directly. Moreover, the contractor and an administrator assert that the report supplied the Administration with reliable information about the amount of time that Title I students spent in instruction, and this information was used to justify requiring a minimum of 5 hours of instruction per week. However, the House Report indicates that the committee felt this minimum requirement to be too small. This amendment became Section 152 of Public Law 95-561.

One of the administrators we interviewed noted that the intent of Title I legislation had already been to supplement, not supplant state funds. However, he noted that the Neglected and Delinquent Program still found violations from time to time. The amendment had the effect of emphasizing this requirement for state institutions in particular.

In both the Department of Education and the Congress, discussion continues about the means of improving services to students in the institutions. For example, instruction of students incarcerated in adult institutions is only provided at the discretion of the states. Time in instruction has been found to be critical for achievement, and so there is much debate over the means of increasing instruction for Title I students in institutions. Senator Pell has introduced a bill for education of students in corrections institutions and, according to the program manager, Congressional staff call the manager for information bearing on the bill.

The Office of Program Evaluation of the Department of Education has recently issued a request for proposals for an evaluability assessment that will cover nine substantive areas of education program management. One of these, according to the SDC contractor, involves the Neglected and Delinquent Program. The contractor believes that the RFP emerged as a direct consequence of the SDC evaluation's having identified problems with the program.

The program manager noted that state and local educators have taken the results of the evaluation back to their agencies and institutions in order to show these agencies that there are problems with existing arrangements and to argue for improvements. We interviewed one of these educators, the Director of Education of a school for delinquent youths. He was one of the program consultants for the SDC evaluation. He brought information about preliminary findings back to his school, and gradually the teachers in the institution became sensitized to similar problems in their own school.

So for example, the curriculum committee made some major changes in the courses offered by the school in the fall of 1979. The SDC report had reinforced for them what the teachers had been saying about the school. The changes introduced were some mini-courses in three major areas. Some of these dealt with survival skills that the students would need on their release, such as information about family and peer relations. Others dealt with life skills, such as functional writing and consumer information. The third major area consisted of an orientation to industrial arts.

The Director of Education argues that the SDC report had enabled the school to argue cogently before the state legislators for the changes they had made. They were able to go to the central corrections office and speed up changes that might have taken place anyway, but would have taken at least a year longer. Because they had data, in addition to their own experience with the school, they were able to refute those who opposed their changes.

Several of the people we interviewed noted that the study has been useful in part because of the good relations among the contractor, the project monitor, and the program manager. In fact, the program manager welcomed the evaluation because it brought attention to problems which therefore had a higher likelihood of being addressed. That they may be addressed is evident from legislative debates over further action, and dissemination of the results to state and local institutions.

In developing this case study we interviewed Janice Anderson, Project Monitor for the evaluation of Neglected and Delinquent Youth Program, Office of Program Evaluation; Theodore Bartell, Principal Investigator for the Systems Development Corporation evaluation; Ellen Balko, Procurement Officer for Office of Program Evaluation contracts; Chris Cross, former Congressional staff member; John Hoyt, Investigator for the Systems Development Corporation evaluation; Pat Mancini, Education Program Specialist for the Neglected and Delinquent Program; Paul Miller, Program Support Branch Chief, Division of Education for the Disadvantaged; and James Wickman, Director of Education of the Lincoln Hills School, Irma, Wisconsin.

References

Education Department, Lincoln Hills School. Program description. Irma, Wisconsin: Lincoln Hills School, June, 1980.

Systems Development Corporation, National evaluation of ESEA Title I programs in state institutions for the neglected or delinquent, Phase I. Santa Monica, California: Systems Development Corporation, 1978.

U.S. House of Representatives, Committee on Education and Labor. Report on the Education Amendments of 1978, H.R. 15. Washington, D.C.: U.S. Government Printing Office, 1978.

Case Study: Use of the Sustaining Effects Study

The Education Amendments of 1974 required the Office of Education to report on the numbers of children who were economically and/or educationally disadvantaged who do and do not receive compensatory education services. The Sustaining Effects Study, carried out by Systems Development Corporation with Decima Research Corporation as a subcontractor, supplied this required information. They sampled over five thousand schools, surveying each principal in order to obtain school characteristics, percentage of poor readers, and source of compensatory education funds. They interviewed the families of a subsample of 15,000 children to determine their economic status.

The comparison of economically and educationally disadvantaged children was used extensively in the debate between Congressman Perkins and former Congressman Quie over the criterion for inclusion of Children in Title I. Both Congressmen could use data from the Sustaining Effects Study to support their positions. For example, the Study showed that 39% of low income/low achieving students were being served by Title I, so that Congressman Perkins could argue that the focus of Title I should continue to be the poor. On the other hand, Congressman Quie could argue that because only 40% of low achieving, non-poor students were receiving any type of compensatory education services, the program should be expanded to include the non-poor. Quie could argue that Title I money was going to relatively few attendance areas. Perkins could argue that there was a relationship between the number of poor children in a school and the number of low achievers, and that Title I funds made a greater relative contribution to poor districts. Data from the study were cited supporting Quie's view in the House report (p. 20).

Chris Cross, a former staff member for Congressman Quie, notes that the utility of the Sustaining Effects study was mixed from Quie's point of view, in that it did supply ammunition to both sides. Administrators of Title I that we interviewed noted that, at least the data informed the debate. Moreover, the Sustaining Effects Study and the NIE Compensatory Education Study at least allowed the Office of Education to cite facts to the Committee, rather than their feelings or judgments. These administrators noted that the debate over the scope of Title I will probably continue at one level or another for quite a while.

Several respondents noted that during the course of hearings in 1977 and 1978, Congressional staff requested informally that about 10 special analyses be performed. These requests were channeled through the Congressional Research Service to the project monitor at the Office of Evaluation and Dissemination, who would then request that Systems Development Corporation perform the analyses. These special analyses consisted of projections of the consequences of changing the formula allocations for school districts. It is not clear that these analyses influenced the eventual allocation formula. People have forgotten this information over the two years since the passage of the amendments.

The findings on the numbers of children receiving services did, however, cause a number of activities within Title I management. Some 6% of nonpoor, non-low achieving students received Title I services, according to one analysis, while 48% of low income, low achieving students were receiving no compensatory education services from any source. The causes of these disparities are often beyond the control of the federal program. However, according to one of our respondents, this information has provided the impetus and the resources for Title I to devote more attention to student selection. New technical assistance is being offered to districts to enable them to better select students. This assistance is being offered by the Title I Technical Assistance Centers, as well as central office personnel.

Since the Amendments of 1978, the Sustaining Effects Study has been used in budget allocations in two ways. First, according to two of our respondents, the data helped to justify budgets for Title I for fiscal years 1978 through 1981. The data were also used in Administration responses to questions posed by Congressmen during the appropriations hearings. Secondly, the data were used to check on the accuracy of the Title I administrators' estimates of the number of children served by the program. The last data that had addressed this question had been collected in the late 1960's. From the Sustaining Effects Study, program managers were able to show that their estimates were correct.

One program analyst said that she had used the Sustaining Effects Study when she was working on a study of overlap between Title I and handicapped services under PL 94-142. The project monitor confirms this. Because there seemed to be little overlap (a finding confirmed by the GAO), plans to deal with overlap have been cancelled. In the absence of this information, however, needless activity might have ensued.

An analyst for the Assistant Secretary for Planning and Budget said that he was using data from the Sustaining Effects Study to help the Office of Civil Rights formulate regulations for bilingual education. The Sustaining Effects Study is the only good source of information on the numbers of students involved, and therefore the cost of serving them.

This same analyst said that in the near future he would use Sustaining Effects Study data to examine the adequacy of Title I regulations for targetting services to elligible students, for examining the relative effects of home and school on achievement, and for understanding the effects of TV on achievement.

Some respondents said that the Sustaining Effects Study data were difficult to use or to understand. The tables were cryptic, the rows and columns did not necessarily add up, and outside assistance therefore became necessary. One budget analyst requested assistance from the project monitor, who clarified this information for him. The analyst examing the overlap question sought assistance but was not satisfied even so. This situation may be remedied in the final report, which is to be completed shortly.

The analyst for the Assitant Secretary for Planning and Budget said that because the Sustaining Effects Study's budget was cut by Congress, data were not collected that would have been important for questions he was addressing. The Systems Development Corporation and the Office of Evaluation and Dissemination made the decision that they ought to retain those elements of the sample that were of interest to compensatory education. Because of this decision and because of attrition, the sample was no longer a representative one at the end of 3 years. For the purposes of this analyst, the losses of information were considerable. For the purposes of studying compensatory education, the decision was unfortunate but it might be argued that it was necessary.

Chris Cross has said that Congress was concerned that at the end of the Sustaining Effects Study, the Title I program would be greatly changed, and the relevance of the data therefore limited. He points out that the program has indeed changed a great deal. The Office of Program Evaluation has argued that in reality compensatory education programs do not change very much, in spite of changes in policy. The effect of Congress' action, however, was to greatly reduce the generalizability and detail of the information the study could provide.

In developing this case study we interviewed Janice Anderson, the current Project Monitor for the Sustaining Effects Study, Office of Program Evaluation; Keith Baker, Social Science Analyst for the Assistant Secretary for Planning and Budget; Beatrice Berman, Program Analyst for the Assistant Secretary for Planning and Budget; Vincent Breglio, Executive Vice President, Decima Research Corporation; Launor Carter, Vice President, Systems Development Corporation; Chris Cross, former staff member to Congressman Albert Quie; James Hubbard, Program Analyst for the Assistant Secretary for Planning and Budget; William Lobosco, Education Specialist for Title I; Thomas McNamara of the Office of the Assistant Secretary for Planning and Budget; Paul Miller, Program Support Branch Chief, Division of Education for the Disadvantaged; and George Mayeske, former Project Monitor for the Sustaining Effects Study, Office of Evaluation and Dissemination.

References

U.S. House of Representatives, Committee on Education and Labor. Report on the Education Amendments of 1978, H.R. 15. Washington, DC: U.S. Government Printing Office, 1978.

U.S. Office of Education, Office of Evaluation and Dissemination. Executive summary: Providing compensatory services. Washington, DC: USOE, 1979.

Case Study: Use of an Evaluation of ESAA Nonprofit Organizations

The Emergency School Assistance Act of 1972 supports nonprofit organizations in communities to assist in implementing school desegregation plans. The Rand Corporation studied the effectiveness of these nonprofit organizations and contrasted them with community organizations not funded by ESAA. Rand identified several problems relating to their effectiveness that have led to alterations in the program. Jesse Jordan, Director of Program Operations for the Equal Educational Opportunity Program, notes that, "As a result of this evaluation, the program was changed to the extent that it is really a different program now."

Rand found that the activities of nonprofit organizations were not well coordinated with desegregation activities of the districts. Rand also identified characteristics of effective, as opposed to ineffective, nonprofit organizations, including the use of citizen action strategies such as formation of coalitions with community groups to promote desegregation. OE was funding relatively few such nonprofit organizations and had no means, before the report, of assessing what made for an effective nonprofit organization.

According to Jordan, these findings contributed to Congress' decision to remove the funding of such organizations from the state apportionment for ESAA. Language of Section 608 was changed, in the Education Amendments of 1978, to fund nonprofit organizations through national competitive grants instead. Although the evaluation is not cited in the House report for this particular change, it is cited shortly after in another context. The change to competitive national grants allows OE to fund those nonprofit organizations showing promise of being effective.

New regulations dealing with ESAA nonprofit organizations reflect Rand's assessment of what factors produce an effective organization. All those people interviewed for this case study agree on this. Moreover, the Rand study is cited in response to commentary on the proposed regulations in the Federal Register for April 11, 1980:

> One commenter asked why experience with other community organizations was considered a relevant criterion [for funding applications]... The Rand study of the NPO program indicated that organizations that achieved the greatest impact in promoting desegregation were those that utilized citizen action strategies such as informing the public on desegregation issues or forming coalitions with other community organizations. This finding was taken into consideration in developing the criterion. (p. 25029)

Criteria for applications that reflect the Rand findings include: involvement of community members in the project; sensitivity to the community and population to be served; and "experience in working effectively with community organizations, especially on matters related to school desegregation and race relations" (Sections 185.129 and 185.130).

The Rand study also found that the close relationship between districts and nonprofit organizations often resulted in the nonprofit organizations' providing compensatory education services. In the House report for the Education Amendments of 1978, this report is cited as showing that "two-thirds of ESAA funds are spent on instruction in basic skills." (p. 96) The Committee decided that the emphasis of the nonprofit organizations should be desegregation, not instruction, and therefore revised the list of activities to be funded to focus directly on desegregation. The language of Section 608 of PI 95-561 was changed to this effect.

The Rand study was also used as justification for subsequent changes in regulation dealing with the prohibition of compensatory education using organization funds. In response to comments on the proposed regulations, it is noted:

> An evaluation of the NPO program by the Rand Corporation...revealed that these activities undermine the effectiveness of the NPO in facilitating school desegregation. Moreover, the legislative history of amendments to the Act made by Pub. L. 95-561 calls into question the emphasis given to these activities in the past and indicates a Congressional preference for activities more closely related to the desegregation process (*Federal Register*, April 11, 1980, p. 25028).

In 1979, Congress cut the budget for nonprofit organizations by two-thirds, an action not sought by the Administration. All those interviewed for this case study agreed that the justification for the cuts was the Rand study. The Senate report on appropriations for FY 1980 cites "an HEW study" reporting that the majority of organizations were not actively involved in desegregation. The budget for the program was cut by $10 million, in order that OE would fund only effective organizations.

In developing this case study we interviewed Lawrence Bussey, Special Assistant to the Deputy Assistant Secretary for Equal Educational Opportunity; Jesse Jordan, Director of Program Operations, Equal Educational Opportunity Program; and Robert York, Project Monitor for the evaluation of ESAA Nonprofit Organizations.

References

Federal Register, April 11, 1980. Department of Health, Education and Welfare, Office of Education, Emergency School Aid.

Public Law 95-561, The Education Amendments of 1978. Washington, D.C.: U.S. Government Printing Office, 1978.

Rand Corporation. *Evaluation of the Emergency School Aid Act Non-profit Organization Program*. Santa Monica, California: Rand Corporation, 1978.

U.S. House of Representatives, Committee on Education and Labor. Report on The Education Amendments of 1978, H.R. 15. Washington, D.C.: U.S. Government Printing Office, 1978.

U.S. Senate Committee on Appropriations. Report on the Departments of Labor and Health, Education, and Welfare, and related agencies appropriation bill, 1980. Washington, D.C.: U.S. Government Printing Office, 1979.

Case Study: Use of an Evaluation of Magnet Schools

In 1979 Abt Associates completed an evaluation of OE's magnet schools as a tool for desegregation. Magnet schools have specialized curricula and other resources that are intended to attract students to the school and so facilitate voluntary desegregation of the community. Abt Associates interviewed administrators, teachers, parents, and community groups involved in desegregation in 18 school districts containing magnet schools. They assessed conditions under which magnet schools are likely to be effective in desegregation.

The evaluation noted that magnet schools are successful in achieving desegregation within the school and are associated with desegregation in the district as a whole. Moreover, Abt noted that the communities had more positive attitudes toward desegregation after experience with magnet schools. However, the report cautioned readers that the findings on district-wide desegregation were qualified by several problems, and that the findings on attitudes were "at best, suggestive" (p. 11).

Nevertheless, these positive findings were cited. Prior to the evaluation, there was a general realization that magnet schools were an untested concept. Because the evaluation was generally favorable, several of the people interviewed for this case study believe that the report put an end to debates over the efficacy of magnet schools. This belief is also reflected in the House Report on the Education Amendments of 1978:

> ...an evaluation of the ESAA magnet schools by Abt Associates concluded that in every site visited, people felt these schools had a positive effect on community attitudes. (p. 93)

The Senate Committee on Appropriations also commented on the positive findings with respect to magnet schools in its 1979 report:

> The Committee has included $50,000,000 for magnet schools; this is an increase of...$25,000,000 over the 1979 appropriation. The magnet school is one of the most effective tools for voluntary desegregation. A recent evaluation shows that magnet schools are an effective tool in helping to improve community attitudes toward schools. (p. 107)

However, the Abt evaluation did indicate some problems with the program and areas in which its effectiveness could be increased. The evaluators discovered that OE was awarding magnet school funds to districts with poor records of desegregation, and to districts that had little need for desegregation assistance. Moreover, they discovered that magnet schools were more effective as part of a comprehensive desegregation plan. According to Monika Harrison, Special Assistant to the Deputy Assistant Secretary for Equal Educational Opportunity, these findings confirmed what the Administration had already suspected.

According to Jesse Jordan, Director of Program Operations of the Equal Educational Opportunity Program, the Department of Education is forbidden by law to require that magnet schools be part of an overall desegregation plan. The Abt finding that magnet schools were more effective in such an overall plan was brought to the attention of the House of Representatives Committee on Education and Labor in hearings in March of 1980 (transcripts were not available as of this writing). According to Jordan, Congress will probably not change the law, because in spite of the Abt Associates finding, it still hopes for results from the magnet school concept alone.

The Department of Education was able to revise regulations to take into account the school district's desegregation record, however. Jesse Jordan had primary responsibility for writing new regulations for the ESAA programs. Jordan said that he used the Abt findings in changing the rankings of applications for magnet school funds. The new regulations give priority to those applications from districts that have achieved reductions in the isolation of minority students. The relevant regulation is section 185.104, published in the Federal Register for May 16, 1980.

Funding for the magnet school concept has been rising rapidly at least in part because the program is popular with Congress. According to Monika Harrison, the Administration argued that funding for the program should be cut in light of the Abt finding that the program was growing out of proportion to actual desegregation activities. According to Harrison, who assisted in the preparation of budgets, the Administration requested reduced funding for the program for FY 1981 and a recision in funding for FY 1980 (transcripts of hearings not available as of this writing). Jordan agrees that the Abt findings provided part of the Administration's rationale for the cuts, but it is not clear whether the report was cited. Jordan notes that Congress did not agree to the cuts in funding requested by the Administration. On July 3, 1980, the Congress did reduce the budget for magnet schools by $6 million, but Jordan believes that this was part of an overall budget reduction and was not intended to cut magnet school funding *per se*.

Abt Associates noted several reasons to change the regulation that no more than 50 percent of magnet schools be minority students. The requirement helped discriminate against districts with large minority populations; it demeaned minorities who already felt that the program benefitted whites; and more generally, the focus of the program was felt to be the district, not the school. Both Jesse Jordan and David Lerch, Program Manager for Magnet Schools, said that these findings were used in changing the regulations. The new regulations allow more flexibility by focusing on the district enrollment as a whole, not on the school. The relevant section is 185.101 of the regulations as published in the Federal Register for May 16, 1980.

In developing this case study, we interviewed Monika Harrison, Special Assistant to the Deputy Assistant Secretary for Equal Educational Opportunity; Jesse Jordan, Director of Program Operations, Equal Educational Opportunity Program; David Lerch, Special Projects Branch Chief, Equal Educational Opportunity Program; Eugene Royster, Principal Investigator of the Abt

Associates Magnet Schools Evaluation; and Robert York, Project Monitor for the Magnet Schools Evaluation, Office of Program Evaluation, Department of Education.

References

Abt Associates, Inc. Study of the Emergency School Aid Act Magnet School Program. Cambridge. Massachusetts: Abt Associates, 1979.

Federal Register, May 16, 1980. Department of Education, Emergency School Aid. Washington, D.C.: U.S. Government Printing Office, 1980.

U.S. House of Representatives, Committee on Education and Labor. Report on the Education Amendments of 1978, H.R. 15. Washington, D.C.: U.S. Government Printing Office, 1978.

U.S. Senate, Committee on Appropriations. Report on the Departments of Labor and Health, Education, and Welfare, and related agencies appropriation bill, 1980. Washington, D.C.: U.S. Government Printing Office, 1979.

Case Study: Use of the Survey of ESAA-TV Viewership

The Office of Education has funded the development of public television series to assist in reaching the goals of the Emergency School Assistance Act. These television programs were aimed at reaching minority group children of school age. A survey of the viewership of these programs was conducted by Applied Management Sciences. Less than one third of the children interviewed recalled having seen even one of the ESAA-TV series. Less than 10% of the children preferred the ESAA-TV programs over other programs being shown at the same time. The evaluators note that this problem is typical of "educational" TV programs; compared to other such public television programs, ESAA-TV had achieved reasonable viewership.

Both the program people we interviewed and the project monitor agreed that federal managers were already aware that there were problems with viewership at the time the survey was completed. Some management changes were already under way at the time the survey was completed, while others were being considered and were later implemented. For example, a subcontract had been let to improve viewership. A contract was let after completion of the survey for converting the programs to school use, rather than home use.

One reason for low viewership was that the program had concentrated on development of series, rather than promotion of them, according to the former director of ESAA-TV. Only gradually did the problems in viewership emerge. The project monitor, when asked why the survey was conducted, said that it is difficult to determine exactly when managers had realized there were problems. He said that the study was closely coordinated with the managers.

Because changes in the program geared to increasing viewership are under way, the data collected in the survey may become obsolete. The principal investigator of a study of ESAA-TV administration, funding, and local stations carrying the programs, noted that the viewership situation has changed radically because of these changes. Therefore, her group may perform secondary analysis of the survey data to supplement their own report, but recognize that it is dated.

Although the report was intended primarily for the managers of the program, it was used by the Department of Education, according to Monika Harrison, Special Assistant to the Deputy Assistant for Equal Educational Opportunity, in budget requests. Harrison assisted in preparation of the budget proposals for the program in 1979. Because the data showed the necessity for promotion Harrison believes that they allowed the Administration to argue cogently for funds to promote the series, in hearings before Congress and OMB. However, Congressional hearings make no mention of the study. Moreover, it is distinctly possible that this use occurred informally, within the administration.

Jesse Jordan, Director of Program Operations of the Equal Educational

Opportunity Program, notes a use of the survey in regulations, although again the information corresponded to what was already known about the program. Jordan says that he used the survey results in writing new regulations that changed the funding for producing the television series from grants to contracts. This statement is corroborated by others. This arrangement gives the Department of Education more control over what kinds of programs will be produced, thus enhancing viewership. The relevant sections of the new regulations are 185.150 through 185.155 as published in the Federal Register for May 16, 1980. In response to commentary on the proposed regulations, the following rationale is given:

> These requirements are based on program experience in managing the educational television program authorized by the statute as originally enacted, are necessary to implement the statutory mandate that "programs ... be made reasonably available for transmission, free of charge, and shall not be transmitted under commercial sponsorship."(p. 32654)

In developing this case study we interviewed David Berkman, Assistant Dean in Charge of Telecommunications, Newhouse School of Public Communications and former Director of the ESAA-TV program; Malcolm Davis, Director, Division of Educational Technology and director of the program; Monika Harrison, Special Assistant to the Deputy Assistant Secretary for Equal Educational Opportunity; Jesse Jordan, Director of Program Operations, Equal Educational Opportunity Program; Arthur Kirschenbaum, Project Monitor for the ESAA-TV survey, Office of Program Evaluation; Bernadette Nelson, Principal Investigator for Part II of the ESAA-TV study, currently being conducted by Abt Associates; and Anne Kuchak, Vice President, Applied Management Sciences.

References

Applied Management Sciences. A survey of viewership of television series sponsored by ESAA legislation. Silver Spring, Maryland: Applied Management Sciences, 1978.

Federal Register, May 16, 1980. Department of Education, Emergency School Aid. Washington, DC: U.S. Government Printing Office, 1980.

Case Study: Use of an Evaluation of Implementation of Project Information Packages

This study, conducted by American Institutes for Research, assessed that innovative education projects are implemented at sites other than those in which the projects were developed. The evaluators found that the dissemination strategy and the instructional packages created to facilitate implementation were moderately useful. However, the newly implemented program differed notably from the original innovation in many of the sites that were studied.

Evaluation results were presented in a briefing session for federal managers who had direct responsibility for the innovative projects and their dissemination. The briefing was aimed at challenging assumptions people had held about the program. One assumption was that the whole program would be adopted rather than parts. Adopters tended to take from an innovation the elements that appealed to them. The briefing also challenged the idea that adopters would eagerly seek advice from the developer. The evaluators concluded that to obtain adoptions that did not differ notably from the original, incentives must be created.

The instructional packages developed to facilitate program implementation, Project Implementation Packages, originally represented a management experiment, according to several of the people we interviewed. They were developed in order to determine whether they assisted in implementation. This evaluation and other research convinced some managers that other actions were necessary. Others maintain that they believed all along that other actions were necessary and the evaluation strengthened their belief.

Several people interviewed agreed that subsequent changes in the National Diffusion Network, the organizational vehicle for making the innovations available, were at least in part attributable to this study. First, developers of the innovative projects were funded to give individual assistance to project adopters. Second, OE contracted to provide developers with technical assistance to more effectively disseminate their projects, because the developer of an innovation is not always able to communicate this information effectively.

Another modification of the program reflects concern over the extent to which adoptions of the projects are similar to the original. The Office of Dissemination and Replication is currently attempting to determine criteria for the quality of adoptions. New regulations published in the Federal Register for April 21, 1980, require that every four years, developers present evaluation evidence of the quality of adoptions in order to be recertified. The relevant sections of the regulations are 193.12 and 193.16.

Several managers in the Office of Dissemination and Replication noted that, in the early years of the National Diffusion Network, primary attention had been given to increasing the acceptance of the program and gaining as many adoptions as possible. The evaluations of the Project Information Packages were part of a watershed for the Network. They are now trying to change the presumptions of adopting sites and to persuade them to accept the idea of evaluation as an integral part of the projects. A contract to produce an evaluation guide to assess the quality of adoptions has been made with the Center for the Study of Evaluation at UCLA and has been under development for a year.

The CEIS contact person for the study of Project Information Packages was contacted for this case study. In the opinion of this independent observer, the study was well done technically given the constraints of field research. However, this individual noted that timeliness of the report had been something of a problem. The negotiations between AIR and OE over content and methodology were to some extent responsible. As a reviewer, he wanted more data to support the descriptive conclusions of the study. AIR followed his suggestions on redrafts of the report.

In addition, this individual noted that OE, in conducting the series of Project Information Packages studies, had not made full use of existing literature on the subject of adoption of innovations. For example, he cited one suggestion of an evaluator, that implementation of a new project might be broken down into project elements, rather than whole programs. However, this idea had been current in the literature for many years. In fairness to OE, however, he noted that one has to try such ideas to find out if they will work.

In developing this case study we interviewed Will Ashmore, Program Evaluation and Planning Consultant to the Wisconsin Department of Public Instruction; Anne Bezdek of the Office of Program Evaluation; Judy Burnes of the Office of Program Evaluation and project monitor for the evaluation; Peggy Campeau, principal investigator for the AIR study; Andrew Lebby of the Office of Dissemination and Replication; Louis Walker of the Office of Dissemination and Replication; and Lee Wickline, Director of the Office of Dissemination and Replication.

References

Campeau, P.L. et al., Final report: Evaluation of Project Information Package Dissemination and Implementation. Palo Alto, California: American Institutes for Research, 1979.

Federal Register, April 21, 1980. Department of Health Education, and Welfare, Part V, Office of Education. National Diffusion Network Program: Final Regulations Governing the Award of Grants. Washington, D.C.: U.S. Government Printing Office, 1980.

U.S. Office of Education, Office of Evaluation and Dissemination. Impact of evaluation activities on OE programs. Mimeo.

U.S. Office of Education, Office of Evaluation and Dissemination. National Diffusion Network Policies: Adoption Definition. January 31, 1979, mimeo.

Case Study: Use of a Search for Exemplary Career Education Projects

The Office of Education contracted with American Institutes for Research to conduct a search for effective projects in career education. A second purpose of the contract was to identify ways of improving career education evaluations. AIR asked knowledgeable individuals to nominate outstanding career education projects. Ten of these passed AIR's criteria of evidence of effectiveness. These projects were visited in order to verify the information they had provided, and in order to collect additional data. AIR then prepared documentation to present each project for submission to the Joint Dissemination Review Panel (JDRP) of the Office of Education and NIE.

The JDRP reviewed all ten projects and judged seven as having proven their effectiveness. These seven projects thus became eligible for grants as developer/demonstrators in the National Diffusion Network. Six of the projects did apply for such grants.

Evidence of use of the projects consists of the numbers of adoptions of the projects by other sites, and the number of inquiries received by the developer/demonstrators. We were able to contact four of these projects for information. The remaining two were closed for the summer. We also asked project personnel their opinion of AIR's technical assistance in helping them pass the JDRP.

Project CAP of Greenland, Arkansas provides career awareness information through the regular school curriculum. It was approved for dissemination for ages 1 through 8. They have been funded for dissemination since October of 1979, or about 9 months as of this writing. Thirty-five sites in 8 states have adopted their project. They receive an average of 5 or 6 inquiries a week. If they had not been passed by the JDRP, 5 sites would nevertheless have adopted their project. The director of the project said, "AIR was helpful. If we develop new projects we will ask them to help us evaluate them." She mentioned that there were already plans afoot to submit the project, but without AIR's assistance, "it would have died in the water." For this project, therefore, 30 adoptions in 9 months can be attributed to AIR.

Career Development Programs, of Akron, Ohio, also uses career education activities as part of the ongoing curriculum. The program is approved for grades K through 10. Two years after approval by the JDRP, they have 10 agreements to adopt the project that are in various stages of implementation. They have sent materials used by the project to 150 school districts in 32 states. While using project materials does not constitute an adoption, it may constitute use of project information at some level. They have made formal presentations to 322 districts who inquired about the project. The director had a positive attitude toward the AIR assistance he received.

Project CERES (Career Education Responsive to Every Student) of Ceres, California develops students' decision-making skills and attitudes toward work within the regular school curriculum. Funded since October of 1979 they have had 17 adoptions and approximately 4,500 inquiries by letter and phone. The director said, "Honestly, we would not have made it through the JDRP without AIR. We would never have actively solicited adoptions. We wouldn't have even tried. The technical assistance was worth it." Here are 17 more adoptions that would not have occurred without AIR's work.

Project MATCH (Matching Attitudes and Talents to Career Horizons) of Ontario California infuses career education into the regular school curriculum for grades K through 8, with a component for staff development and self-evaluation. Two years after JDRP approval, they have 24 formal adoptions, plus 20 adoption agreements at various stages of implementation. Eighty sites have purchased the materials for their program. They receive about 1000 inquiries per year. The director said that the AIR activites provided them with a catalyst for submitting their project to the JDRP, something they might not otherwise have taken the time to do. He said that AIR had a good track record in screening projects for effectiveness. From this project therefore we have 44 adoptions that in all probability would not have occurred without AIR's assistance.

Another product of the AIR report was a monograph, published for the Office of Career Education, on getting JDRP approval for career education projects. This monograph is in the process of publication, so that people have not had a chance to use it yet.

A program analyst for the Office of Dissemination and Replications, which administers the National Diffusion Network, said that at first the AIR study had suffered from a typical problem of technical assistance to projects: a feeling of non-involvement on the part of the developers. However, the developers did begin to feel involved as the study progressed. He believes that the projects function reasonably well. A factor that has greatly assisted the dissemination of the projects, according to this analyst, has been the transfer of funds for dissemination from the Office of Career Education to his own office.

In developing this case study we interviewed Darvel Allred, Project Specialist, Ontario-Montclair School District, Ontario California; Jack Hamilton, Principal Investigator for the AIR study of exemplary career education projects; Nancy Keenan of the Office of Career Education; Andrew Lebby of the Office of Dissemination and Replications; Jeanne Leffler, Director, Project CAP, Greenland Arkansas; Virginia Lish, Curriculum Specialist, Ceres School District, Ceres, California; Seymour Rubak, Executive Secretary for the Joint Dissemination Review Panel and Office of Program Evaluation; and Nick Topougis, Director of Career Education Programs, Akron, Ohio.

References

Hamilton, J.A. and Mitchell, A.M. Final technical report: Identification of evaluated, exemplary activities in career education (K-12). Palo Alto, California: American Institutes for Research, 1978.

U.S. Office of Education, Education programs that work. San Francisco: Far West Laboratory, 1979.

U.S. Office of Education, Office of Career Education. How to get JDRP approval of career education projects. Mimeo, 1980.

Case Study: Use of a Survey of Campus-based Aid

A survey of the beneficiaries of four postsecondary student aid programs was conducted by Applied Management Sciences. These programs were: Basic Educational Opportunity Grants (BEOG); Supplemental Educational Opportunity Grants; National Direct Student Loans; and College Work-study. Part one of this study examined existing records and interviewed managers and the federal and regional levels. Part two involved the collection of primary data.

According to several people we interviewed, this study was useful largely because it provided analyses of students benefitting from the program by race, sex, and other background characteristics essential to policy analysis. Other surveys and program data have produced information on characteristics of those served, but the data have been limited in their usefulness by the fact that they involved only one program, or only certain institutions. The current survey permits an examination of funding of all kinds at the individual student level.

The Director of Quality Assurance of the Office of Student Financial Assistance believes that the survey is one of the better studies undertaken by the postsecondary education evaluation group in the Office of Program Evaluation. He notes that its usefulness depends heavily on its comprehensiveness. He maintains that he needs this comprehensive information for long range planning, for determining whether the funds are going to the people that policy intends, and for answering questions about funds that go to particular groups or particular institutions.

The data on student characteristics was shared with several private groups interested in postsecondary education. One of these was the American Council on Education. A policy analyst for ACE said that she produced some statistical analyses of the survey data which were used by her group in developing their own policy position on campus-based aid. In particular, ACE was concerned with the half-cost provision for funding under the Basic Educational Opportunity Grants. Under this limitation, a student's Basic Grant may not exceed 50% of the cost of this or her attendance. The survey gave ACE support for their contention that this provision was hitting low income students the hardest. The position taken by ACE is important because, together with other groups interested in postsecondary education, they produced for the House Sub committee on Postsecondary Education a proposal to alter BEOG which was, according to the House Committee report, "substantially incorporated into H.R. 5192 [the House bill]". (p. 18). This proposal represented a compromise among these groups which over the course of several years raised the amount of money a student could receive under BEOG. In addition the percentage of the cost of education that could be covered by BEOG was to increase to a maximum of 75% in 1985. As of this writing, the proposed changes in BEOG have not yet become law.

The former project monitor for the Survey said that he had responded to several informal requests from Congressional staff for analyses of the data by race and sex of the participants. At the time, this project monitor was assigned part time to the Deputy Commissioner's Office to develop background information for the reauthorization hearings on higher education.

In developing this case study, we interviewed Ernst Becker, Director of Quality Assurance, Office of Student Financial Assistance; Dr. Salvatore Corallo, Director of the Division of Postsecondary Programs, Office of Program Evaluation; Alexander Ratnofsky, former Project Monitor for the study of Campus Based Aid; and Patricia Smith, Associate Director of Policy Analysis Services, American Council on Education. Time did not permit obtaining corroborative testimony from Congressional support staff.

References

Applied Management Sciences. Stage I final report for the study of program management in the Campus-based and Basic Grant Programs. Silver Spring, Maryland: Applied Management Sciences, Inc., 1978.

U.S. House of Representatives, Committee on Education and Labor. Report on the Education Amendments of 1980, H.R. 5192. Washington, D.C.: U.S. Government Printing Office, 1980.

Case Study: Use of the Vocational Education Equity Study

The Education Amendments of 1976 mandated a study of the extent of sex discrimination and sex stereotyping of vocational education programs supported under the Vocational Education Act and an assessment of progress in overcoming such discrimination. American Institutes for Research carried out this study in the spring of 1978. They visited state Vocational Education Offices and a large sample of high schools, vocational schools and technological institutes, and community/junior colleges. The study showed that students continued to be concentrated in classes stereotyped as "appropriate" for their sex. The teachers of vocational education were overwhelmingly concentrated in classes stereotyped as "appropriate" for them to teach.

The study evoked immediate interest from public groups of all kinds, including groups that monitor educational equity, employers, labor unions, state and local commissions on the status of women, vocational education colleges, state and local education associations, and even the Army White Sands Missle Range. The staff member of the public relations office of the Department of Education assigned to distribute information about the study says that it has generated more interest than any other report she has seen since she has been in public relations. After distributing over 350 copies of the executive summary, the Public Relations Office is on its second printing.

However, use of the information has by and large been confined to that of background information. We called a sample of the groups inquiring about the study. These included the NAACP Legal Defense Fund; ACLU of Georgia; Project Opportunity, a group sponsored jointly by the Center for Women and Work and the Coalition of Labor Union Women, and the National Advisory Council on Women's Educational Equity. All these groups said that the study had provided important background information. However, they had not yet used the information for any specific decisions.

The study may be used heavily for decisions in the near future. Reauthorization hearings for the Vocational Education Act will take place in 1981. Several of the groups with whom we talked are developing position papers and materials for these hearings that cite the Vocational Education Equity Study extensively. These include the National Advisory Council on Women's Educational Equity and the National Advisory Council on Vocational Education. For the moment, there is little political action to be taken.

Some state sex equity coordinators may be making use of the report. For example, the coordinator for the State of Iowa says that the report has been "extremely useful" in giving direction to state equity efforts. For example, the "promising approaches" to the elimination of sex bias that are described in the evaluation have shown Iowa that such projects are feasible, and the state is funding projects of this kind for the first time this year. The state coordinator also distributed the executive summary to provide guidance to the first meeting of the state sex equity council.

Two of our contracts expressed serious reservations about the way that the study was handled. Both the Special Advisor on Women's Issues of the Office of Vocational and Adult Education, and the Executive Director of the National Advisory Council on Women's Educational Equity maintain that the conclusions of the report were misinterpreted. Many people misread the report and believe that it evaluates the sex equity coordinators. However, the legislation authorizing sex equity coordinators was passed in October of 1977, only three months before AIR began to collect evaluation data. In fact, the report cautions readers that fully half of the coordinators were very new to the job. However, this information is easily overlooked. Because the report concludes that as of 1978 vocational education was still sex-stereotyped, it gives the unjust appearance that sex equity coordinators are not being effective.

The Special Advisor to the Office of Vocational and Adult Education mentioned, for example, that one of her colleagues heard the evaluation cited as evidence against sex equity coordinators in a meeting of the National Commission on Employment Policy. According to the Special Advisor, many state coordinators are concerned over such misinterpretations of the report. The Special Advisor noted that the state coordinators would, in fact, welcome an evaluation of their efforts, now that their positions have been in existence for three years.

The Special Advisor and the Executive Director of the National Advisory Council noted other problems related to utilization. The clearance process for the report was exceptionally long. The evaluators finished collecting data in spring of 1978, and the report was only released in February of 1980. The fact that the study has only been available for six months as of this writing may well have affected the degree of use to which it has been put. Dissemination of the report has been hindered by its cost, which is $35 for the complete set of volumes. Although the Special Advisor regards as useful the section of the report on "promising approaches", the Executive Director of the Council cannot recall any of them being adopted by people with whom she is acquainted.

In creating this case study we interviewed Barbara Bitters, Special Advisor on Women's Issues of the Office of Vocational and Adult Education; Beverly Gillette, Non-sexist Vocational Education Consultant (equivalent to a sex-equity coordinator) for the Department of Public Instruction of the State of Iowa; Laurie Harrison, Principal Investigator for the AIR study of Equity in Vocational Education; Charlotte Hoffman of the Public Relations Office; Virginia Looney, Coordinator of the Vocational Education Monitoring Project, ACLU of Georgia; Phyllis McClure of the NAACP Legal Defense Fund; Shirley Robock of Project Opportunity; Dorothy Schuler, Project Monitor, Office of Program Evaluation; and Joy Simonson, Executive Director of the National Advisory Council on Women's Educational Equity.

References

Gilette, B. Letter to members of the Iowa State Sex Equity Advisory Committee, April 15, 1980.

Harrison, L.R., and Dahl, P.R. Executive summary: Vocational education equity study. Palo Alto, California: American Institutes for Research, 1979.

Case Study: Use of an Evaluation of Upward Bound

Research Triangle Institute conducted an evaluation of Upward Bound that followed the progress of students through high school and entry into college. Among the major results were the following: students were more likely to enter college if they had participated in the program for two or three years; of those students not immediately entering college, a much larger percentage of the Upward Bound students entered college within 3 years; and Upward Bound increased the number of minority and poverty students in college. This evaluation is the second follow-up study of an evaluation originally conducted in 1976, but is the first study that contains detailed information about the impact of the program.

All federal agency staff that we interviewed maintained that a major use of the evaluations occurred each time Congress appropriated money for the program. One analyst involved in the preparation of the program budget for the Department of Education noted that the positive evaluations helped show Congress has always appropriated more money for the program than the Administration requests. Moreover, it is likely that the evaluations of Upward Bound have had something to do with these decisions. After being held constant for 4 years, funding for the program has increased steadily since the first evaluation of the program was released in 1976. However, apart from the opinions of the federal agency staffers, we have been unable to establish any direct link between the results of the evaluation and Congressional budget decisions.

The Education Amendments of 1980 will probably change the current eligibility criteria for Upward Bound to 150% of the Orshansky index of poverty. Students with no family tradition of college attendance, regardless of their poverty status, will be allowed to participate as well. The evaluations of Upward Bound probably influenced these decisions indirectly. All students in the program were found in the study to need Upward Bound services, but only three fourths of students in the study met the current poverty criterion. Lobbying groups in post-secondary education were, according to the project monitor, able to use this information as support for their position on changes in Upward Bound that will probably be adopted.

Upward Bound staff are now writing regulations for the program in anticipation of changes in the law. Criteria for funding Upward Bound projects will probably include the prior performance of the projects. Prior performance will be evaluated in terms of: persistence of students in higher education; proportion of students completing high school; and proportion of students entering postsecondary education. These criteria have been demonstrated as measurable by the evaluation, so that the regulations writers view them as feasible. Two of the program people we interviewed confirmed this use of the evaluations.

The first evaluation of Upward Bound was also used in regulation writing, according to the project monitor. This study had found that two or more years of Upward Bound are necessary for the program to make a difference. This finding was used in writing the current regulation that senior in high school may not be admitted to the program as beginning clients.

In developing this case study we interviewed Dennis Carroll, Project Monitor for the evaluation of Upward Bound, Office of Program Evaluation; Salvatorre Corallo, Director, Division of Postsecondary Programs, Office of Program Evaluation; James Herbert, Program Analyst for Higher Education Programs, Office of Planning and Budget; Shelly Laverty, Program Aanlyst, Office of Postsecondary Education; and Velma Montero, staff member, Upward Bound Program. We did not have sufficient time to corroborate this testimony with information from outside the federal agencies, e.g., Congressional support staff, aside from CBO (see the case study on the Congressional Budget Office).

References

Corallo, S.B. and Carroll, C.D. The evolution of USOE evaluation of the "TRIO" Programs. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, May, 1980.

Research Triangle Institute. Evaluation of the Upward Bound Program: A second follow-up. Durham, North Carolina: Research Triangle Institute, 1978.

U.S. House of Representatives, Committee on Appropriations. Report on Departments of Labor and Health, Education and Welfare, and related agencies appropriations bill, 1978. Washington, DC: U.S. Government Printing Office, 1977.

U.S. House of Representatives, Committee on Appropriations. Report on Departments of Labor and Health, Education and Welfare and related agencies appropriations bill, 1979. Washington, DC: U.S. Government Printing Office, 1978.

U.S. House of Representatives, Committee on Appropriations. Report on Departments of Labor and Health, Education and Welfare, and related agencies appropriations bill, 1980. Washington, DC: U.S. Government Printing Office, 1979.

U.S. Office of Education, Office of Evaluation and Dissemination. Annual evaluation report on programs administered by the Office of Education, 1979. Washington, DC: USOE, 1980.

Case Study: Use of an Evaluation of Services in Indian Education

A descriptive study examined projects at over 200 sites that were funded under Part A of the Indian Education Act. Projects were most frequently directed toward: instruction of cultural heritage and native language, remedial reading, improvement of self-concept, and remedial mathematics. Successful implementation of projects was found to be facilitated by the size of the grant, type of objectives, parent involvement and density of Indian population.

According to a planning officer for the Office of Indian Education, the findings of the study provided evidence for several assumptions that the Office had held for a long time. These included the importance of parental involvement to the projects, and the belief that projects will probably not show effects if funding is below a certain level. Because school districts are funded on the basis of the number of Indians attending, low funding levels may interfere with the success of some programs. They are therefore trying to persuade Congressional Committees to increase appropriations to allow for increased per pupil expenditures. This officer did not cite any hearings, however.

This same planning officer found that information on funding patterns was useful because it revealed that school districts with higher concentrations of Indian children have a greater need for services. This is true because Indian land is not taxable. Indian children therefore represent a financial burden for districts that must be redressed through these programs. Judith Baker says that Indian Education did use the survey in their budget submission to OMB and Congress for Fy 1981. It is entitled, "Justification of appropriations estimates for committees on appropriations, FY 1981, Indian Education, Department of Education." The relevant material appears on Page 12 of the document. This occurred about a year ago. She says that the material did help them by providing statistics -- the specifics that showed they were well prepared.

Both the planning officer and the Deputy Assistant Secretary for Indian Education agreed, however, that the study had not been particularly helpful. All the study did, in the words of the Deputy, was "list what was out there." They would both have preferred an impact study or one which showed which types of projects were effective for Indian students. There are currently plans for such a study in the Office of Program Evaluation.

In developing this case study we interviewed Judith Baker, Deputy Assistant Secretary for Indian Education; Emmet Fleming, Project Monitor for the Study of Indian Education, Office of Program Evaluation; Patricia Matthews, Planning Officer, Office of Indian Education; and Thomas Mullowney, Principal Investigator for the Study of Indian Education, Communications Technology Corporation. We did not have the time to corroborate this information using other sources, such as Congressional support staff.

References

U.S. Office of Education, Office of Evaluation and Dissemination. Executive summary: Assessment of programs and projects funded under Title IV of PL 92-318, Indian Education Act, Part A. Washington, D.C.: USOE, 1978

FOOTNOTES

(Chapter 6)

1. The full citation to the published reports of authors cited in this chapter are given in the reference list. Where specification is necessary, we use footnotes to identify particular articles in the reference list.

2. See respectively, David (1978), Emrick et al. (1977) on the AIR Studies, Lyons, et al. (1978) for reports of UCLA's Center for the Study of Evaluation, Alkin et al. (1974), Alkin et al. (1979), Berman and McLaughlin et al. (1977) on the Rand Study and Datta's (1979) critique, and Kennedy et al. (1979) for information on the Huron Institute Study. See Holley (1980) and Webster and Stufflebeam (1980) on the Austin's research.

3. See respectively Emrick et al. (1977), Mitchell (1980), Bissel (1979), and Hope Associates (1979).

4. See the series of Annual Evaluation Reports issued by the U.S. Office Education (1976-1979), U.S. Senate, Committee on Human Resources (1978), and U.S. House of Representatives, Committee on Labor and Education (1978, 1979). The survey results are given in Florio et al. (1980), Fox (1977), and Weinberg (1979). Case studies are reported by Datta (1976) and Millsap (1978).

5. Claims made in this paragraph are based on corroborated interviews.

6. See Mitchell (1980), for instance.

7. See Sproull and Zubrow (1980).

8. Tallmadge (1977b, 1977c).

7. RECOMMENDATIONS

Robert F. Boruch and David S. Cordray

With

Georgine M. Pion and Laura Leviton

7. RECOMMENDATIONS

Recommendations to the Congress are discussed in Section 7.1, those intended for the Department of Education are given in 7.2, along with a brief rationale for each. Section 7.3 contains an extended description of the rationale for selected recommendations.

7.1 RECOMMENDATIONS TO THE CONGRESS

Planning and Executing Evaluations

We recommend that the Congress direct the relevant staff of Congressional committees, GAO, and CBO to meet regularly with evaluation staff of the Department to:

- reach agreement about when particular evaluations are warranted, and the senses in which each evaluation required by law is possible.
- clarify Congressional information needs, quality of evidence required, and planning cycle for each major evaluation required by law.
- identify specific committees and groups as audiences for evaluation results.
- identify the changes in program or understanding which could occur on the basis of alternative findings.

This recommendation hinges partly on the fact that a statutory demand for "evaluation" is ambiguous. The word can imply any activity from journalistic reporting to full-blown field experiments dedicated to estimating the effects of an innovation on children. The involvement of multiple interest groups is often necessary, but complicates matters. At worst, general demands to evaluate obscure the fact that feasibility of evaluation varies enormously and that elaborate evaluation may be unnecessary. Periodic efforts have been made by members of the Congressional staff to assure that production of evaluations coincides with authorization cycles and that Congressional needs are understood. The process is less regular and less orderly than it ought to be.

Statutory Provisions for Evaluation

We recommend that in constructing statutory provisions for evaluation that the Congress:

- specify exactly which questions ought to be addressed and the audiences to whom results should be addressed.
- provide for formal assessment of the evaluability of the relevant program where specification is not possible.
- provide for statistically valid field testing of proposed evaluation requirements where specification is not possible and in-house assessment insufficient.

Though statutes are explicit about routine reporting requirements, references to evaluation often are not specific. The simple requirement to evaluate whether the program meets objectives of the statute is common and vague. Hearings are often not informative. Defining evaluation requirements in terms of the questions which should be addressed is sensible so long as the questions themselves make sense, answering them is feasible, and the answers are likely to be useful. The specification of audiences, especially particular committees or Congressional support agencies, should enhance usefulness. We recognize that explicitness is often not feasible or desirable. Consequently, we suggest formal investigation of evaluability to clarify questions, audiences, and the ways in which results can be used, within a year after enactment of a demand for evaluation.

Evaluator Capabilities

We recommend that

- capabilities be assessed before new statutory evaluation requirements are directed at LEAs and SEAs to determine where resources are adequate to meet the demand
- expansion of training or technical assistance when the demands are notable and capabilities low
- explore the feasibility and desirability of direct contracts programs to capitalize on LEA and SEA capabilities.

The first recommendation stems from conclusions that no real standard for assigning the title "evaluator" exists. Skills required of the evaluator depend heavily on nature of the evaluation demand and on LEA and SEA interest in evaluation. The second recommendation is based on the finding that most LEAs and SEAs need assistance when the demand is high and want it. A small minority of LEAs have strong evaluation units. But these are a major resource and we believe that direct grant opportunities should be expanded to capitalize on them.

Use of and Authority for Better Evaluation Designs

We recommend that the Congress:

- routinely consider pilot testing every new program, variations on existing programs, and program components before they are adopted at the national level, using high quality evaluation designs.
- authorize the Secretary explicitly in each evaluation statute to use high quality designs, especially randomized field experiments, for planning and evaluating new program components, program variations, and new programs.

The rationale for the first recommendation is that higher quality evaluations are more feasible before the program is adopted at the national level. Better designs can be employed and conclusions then are likely to be less ambiguous; political-institutional constraints are likely to be less severe. The introduction of new programs can be staged so that earlier stages are a pilot for later ones. We stress formal tests of new program components and new variations here because such evaluations are not a matter of common practice. We will not learn how to bring about clear, detectable changes without more conscientious tests.

The second recommendation stems from our conclusion, based on this and other research, that better designs must be used if the Congress or the Department wants good estimates of the effects of programs on children. We do not advocate estimating those effects in all cases. The process is complicated under the best of conditions, despite cavalier announcements that the "program was successful because test scores went up" or that it was unsuccessful because they went down. We do advocate explicit authority in statutes for high quality designs, especially randomized experiments to facilitate their use. We believe explicit statutory provision is essential because such designs are the best in principle, and that should be recognized. The authorization should provide for review of the use of these designs.

Critique and Reanalysis of Evaluation Results

We recommend that in statutory requirements for evaluation of major programs, the Congress:

- also require independent, balanced, and competent critique of evaluation results that are material to policy decisions.
- require critique of samples of evaluations submitted by LEAs and SEAs in response to legal requirements
- require that statistical data produced by national evaluations be made available for reanalysis.

By critique here we do not mean adverse commentary. We mean reasoned judgments about whether conclusions drawn from the evaluation are sensible and can inform decisions. The main reason for the recommendation is that such criticism is not routine, but it is essential to enhance credibility of good evaluations, to properly identify poor evaluations as such, and to provide feedback to federal evaluation units, contractors, and grantees about the quality of their work. There is no formal system for competent critique of evaluation reports produced by LEAs and SEAs in response to law, yet many could benefit from criticism.

Use of Evaluation Results

We recommend that the Congress:

- direct staff of relevant committees, the Department, and the GAO to routinely outline which institutions can reasonably be expected to use results of each major evaluation and how such results might be used, during the design stage of every major program evaluation.
- specify exactly which evaluations have been used and why they were used, which have not been used and why they were not used, in authorizations and appropriations committee reports.
- require specific information about changes resulting from evaluation, whenever the law requires SEAs to describe uses of evaluation.
- explore the feasibility of direct competitive grants and contracts programs focused on improving the use of results at the LEA and SEA level.

The first recommendation's origins lie in the absence of any mechanism for planning use at the national level. Simply put, unless specific user groups are identified and some decision options laid out, evaluation results are less likely to be used. Indeed, if there is no clear way to link the evaluation with decisions or considerably better understanding, one can argue that the evaluation shouldn't be done at all. Specifying expectations will also help to make it easier to track utilization and that in turn will help to inform judgments about how evaluation resources could be better allocated. The recommendations to cite useful and useless evaluations in federal reports and to require SEAs and LEAs to record specific changes have the same objectives: understanding use better in the interest of better resource allocation. The suggestion to identify useless evaluation is not an invitation to criticize arbitrarily. We found that some LEAs and SEAs are capable and interested in inventing and testing better ways to use information. The suggestion to expand their opportunities for doing so is based on this.

Standards and Guidelines

Recently developed standards and guidelines for evaluation are not appropriate for incorporation into law. They are sufficiently well developed to recommend that the Congress:

- use such guidelines to understand what can reasonably be expected of evaluations.
- direct that agencies use them as a guide where appropriate to developing criteria for judging evaluation plans submitted by LEAs and SEAs.
- elicit assistance in the interpretation of guidelines from Congressional support agencies, such as GAO, that have been instrumental in their construction.

7.2 RECOMMENDATIONS TO THE DEPARTMENT

Authority for Technical Discussion

We recommend that the Department:

- authorize technical staff of evaluation units to initiate discussion of evaluation plans with pertinent Congressional staff, at their discretion, and refrain from directives which impede direct discussion.

The impetus for the recommendation is simple: Competent evaluators can expect to do a good job only when they have the opportunity to discuss Congress's information needs frequently. Restrictions on the evaluation unit's initiating discussion with Congressional staff of Committees that demand evaluation prevent the job from being done better. We recognize that some restrictions on bureaucratic lobbying for programs are warranted, and that some administrative rules are necessary to keep the process orderly. The lack of clear opportunity to figure out what Congress can use decreases the likelihood that evaluations will be timely, relevant, and credible, and the likelihood that the Congress will find the results useful. Relaxing restrictions will not of course guarantee usefulness.

Planning and Executing Evaluations

We recommend that the Department direct principal evaluation unit staff to meet regularly with relevant staff of committees to:

- negotiate agreement about when particular evaluations are warranted and the senses in which each evaluation required by law is possible.
- clarify Congressional information needs, quality of evidence required, and planning cycle for each major evaluation undertaken by the Department.
- identify specific audiences or groups for evaluation results.
- identify the changes in program or understanding which could occur on the basis of evaluation results.

The rationale for this recommendation is identical to the one offered for a similar recommendation made to Congress. Understanding Congressional information needs is not possible without some regular discussion between technical evaluation staff and Congressional staff. Scarcity of evaluation resources requires better planning and that planning cannot be informed without dialogue among relevant staff.

Tests of New Program Components, Program Variations, and New Programs

We recommend that the Department authorize explicitly the use of high quality evaluation designs, especially randomized experiments, in evaluating new program components, program variations, and new programs, in all regulations which require estimating the effects of innovative changes.

The main justification is that high quality designs lead to far less debatable estimates of effects of programs on children than low quality designs. They are more difficult to execute, and they are more feasible for pilot testing new programs, program variations, and program components, than for estimating the effects of ongoing programs. Explicit authorization would make the importance of good designs plain, and would provide more clear opportunity for competent SEAs and LEAs to exploit them.

Critique and Secondary Analysis of Evaluation Results

We recommend that the Department:

- provide for the independent, balanced, and competent critique of every major evaluation funded by the Department in procurement of evaluations and evaluation policy.
- incorporate into procurement procedures and policy the requirement that all statistical data produced in major program evaluations be documented and stored for secondary analysis.
- create an administrative mechanism for deciding when simultaneous analysis by both the original evaluator and an independent analyst is desirable and feasible, and a mechanism for executing simultaneous independent analyses.

The rationale for this recommendation is identical to the one offered for a similar recommendation to Congress.

Access to and Specification of Reports

We recommend that the Department adopt a policy to:

- adhere to a clearance rule which makes evaluation reports available after a specified period of time.
- specify completely the evaluation documents referred to in the Department's Annual Evaluation Report, the Federal Register, and policy statements.

- include, in every major evaluation report, a list of core recipients of the report, or compiling publicly available lists of core recipients.

The recommendation stems partly from difficulties encountered in obtaining reports under review by the Executive Secretariat and other groups involved in the DHEW clearance process. We also found it difficult to identify reports precisely, when they were cited as evidence of the usefulness of evaluation in developing regulations or policy. The absence of a list of core recipients of reports makes it very difficult to identify potential user groups and to determine if reports were used. The consequence is that what is useless or useful is less verifiable.

The Use of Evaluation Results

We recommend that the Department direct evaluation unit staff or evaluation contractors to:

- provide oral reports regularly as well as written reports on results of major evaluations, and on the uses to which results can be put, to relevant Congressional staff and support agency staff and the program staff within the Department.
- create a system to periodically collect, synthesize, and report specific uses to which evaluations are put.
- improve the Annual Evaluation Report by citing instances of use more specifically.
- direct evaluation staff to meet regularly with Congressional staff to clarify information needs, feasibility of evaluation, audiences for results, and ways in which results can be used to modify programs.

The recommendations are based partly on the finding that use of evaluation results is not tracked conscientiously and the belief that it ought to be tracked to learn how to do evaluations better, and how to better allocate evaluation resources. The rationale for the last recommendation is identical to the one given earlier on planning and executing evaluations.

Implementation

We recommend that the Department:

- routinely require formal measurement of the degree to which program plans match actual operations.
- adjoin research on methods of measuring implementation to the introduction of new programs and program variations.
- create an inexpensive central information system on the time and resources required for full implementation of new programs.

The main reason for the first recommendation is simply that measurement of implmentation of innovations is infrequent. The reason for the second recommendation is that we know little about cheap effective methods of measurement in this arena. The third recommendation stems from the absence of any reasonable empirical guidelines on the time and resources necessary to implement innovative programs.

7.3 RATIONALE

The recommendations are based partly on the Project's findings and on judgments about what needs to be done to improve evaluation practice. We have sought advice and criticism from some Congressional and agency staff, members of the National Academy of Sciences Committee on Program Evaluation, and LEA and SEA staff. But time did not permit any systematic critique.

This description of the rationale is divided into broad topical categories. The pertinent sections of the preceding material are noted.

Planning and Executing Evaluations

The legislative decision to evaluate is complicated by the large number of potential participants: the Congress, the Department of Education's Office of Evaluation, the Congressional Budget Office, and the GAO. The time available to make the decision and to frame specific evaluation questions is variable and often appears to be insufficient. The advice of experts is only sometimes available. The process often leaves ambiguous the type of evaluation which is wanted, the audiences for the evaluation, the probable uses of the evaluation results, and the reasons why an evaluation is wanted. See Chapter 2 for details.

Elements of a More Orderly Process. The actions which appear to be necessary to improve matters include: (a) regular meetings among evaluation staff of the Department and the pertinent Congressional Committees, (b) a planning system which matches evaluations to authorization cycles, (c) information systems which make access to previous work simpler and faster, and (d) identification of groups which can contribute to technical quality of the effort.

Reauthorization Cycles. For the last two years there has been a renewed effort to match the production of evaluations to the reauthorization cycle. We understand from memos and recent activity of the Office of Assistant Secretary for Management that the effort will be sustained. It is imperative to do so if either management or Congress expect evaluations to be used in reauthorization decisions.

Meetings. There is no system of regular meetings among technical staff of the agency or the pertinent Congressional Committees to examine the senses in which a program can be evaluated. We believe such meetings are essential to assuring that formal legislative demands for evaluation are as well informed as possible and that the Department's evaluation unit is equipped to handle them. Ideally, such meetings should be held before the law requiring evaluation is enacted. If that is not possible, formal evaluability assessments should be undertaken as soon as possible after enactment.

Those meetings should focus on the information needs of the Congress, notably on the questions which should be addressed in the evaluation. Apart from enumeration of questions, such meetings should be a vehicle for recording at least some of the reasons for asking the questions, and the audiences to whom answers ought to be addressed. We recognize that decisions based on alternative outcomes of an evaluation cannot always be specified beforehand. But we believe that every effort should be made to do so.

Apart from these fundamental issues, such meetings might address chronic problems. Short term information as well as long term information is often of interest to some audiences for results and because each type demands evaluation resources, some agreement on balance needs to be made explicit, at least occasionally. The important distinctions which influence feasibility of evaluations, new vs. old programs for example, ought to be made emphatically. Because every major evaluation must be tailored, the level of flexibility, what is known and what is not known, ought to be made reasonably clear.

We do not mean to imply that lockstep series of discussion among all relevant staff is warranted or possible. The point is that the absence of regular meetings on Congressional needs virtually guarantees that some needs will not be met. That in turn invites buck-passing and evaluations of lower utility.

Relevant Groups. The groups which should be involved in the process include evaluation staff from the Department of Education's Office of the Assistant Secretary for Management and from the pertinent Congressional Committees. It is sensible to capitalize routinely on support agencies, such as the GAO's new Institute for Program Evaluation and relevant divisions of the Congressional Budget Office and the Congressional Research Service. We do not mean to imply that all agencies need be represented always in lockstep meetings.

Interest Groups and their Role. Interest groups that draft bills which create or modify programs should be urged to provide plans for evaluation of the effects of the proposals. These plans should be routinely reviewed by the Department's evaluation unit if not by a group which includes Congressional staffers and unit staff.

Impediments. There are impediments to any meetings of this sort, of course. On the agency side, for instance, staff have maintained that they have not been free to initiate conversations which would clarify intent of a demand to evaluate, on account of executive policy that restricts discussion. The restrictions are said to have a variety of legitimate origins including preventing agency staff from lobbying directly and independently for pet programs, and to assure that there is at least some orderliness in dealing with the Congress.

For evaluation by units with authority to evaluate, such restrictions are misdirected and inappropriate. No one can conscientiously address a question posed by Congress if the question cannot be discussed directly. We believe that agency policy must recognize the relative independence and discretion of evaluation units.

Impediments on the Congressional staff side appear to be less administrative than physical. To be sure, there are staffers who will participate in no discussion unless directed by a Committee chairman to do so. But they appear to be in the minority. The more general problem is, we are told, time--the sheer difficulty of coordinating meetings so as to be reasonably conventient to both agency staff and Congressional staff. We have not had the time to examine the validity of this complaint. But we find it difficult to believe that it is insurmountable.

There is some vague and episodic reluctance among Congressional staffers to talk to contractors. We do not know how serious the problem is. It does seem sensible to exploit agency staffers as much as possible when Congressional staffers are not familiar with contractors. It seems equally sensible to assure that ingenuous mistrust of contractors does not impede understanding of how the evaluation ought to be conducted.

The impediments are sufficiently formidable to justify some joint action by the Congress and the Department. That action should involve identification of alternative methods of ameliorating the problem and perhaps narrowly focused tests of their feasibilty.

Statutory Provisions for Evaluation

Though some statutes are specific about program reporting, references to "evaluation" in many are very general. The simple requirement to evaluate or to evaluate the effectiveness of the program in meeting the objectives of the statute is frequent.

There is, however, enormous variety in the way individuals at the local, state, and federal level of government interpret the word evaluation in law or elsewhere. It concerns the array of questions which might be addressed in an evaluation, in the approaches one might choose to answer them, and the level of detail at which they might be answered.

More specific statement of the questions which need to be addressed can help to reduce confusion and ambiguity in what is intended by law, and facilitate understanding of scope and probable costs and benefits of the information. See Chapters 2 and 3 for details.

Specification. If the Congress needs to know:

- how many are served and how many are in need,
- what are services and their costs,
- what are the effects of programs on their primary or secondary clients,
- what are the costs and benefits of alternatives,

the Congress should request that information explicitly. That it is feasible to be more specific is clear from the statutes mandating the NIE Compensatory Education Study. That specification is not always sufficient is clear from the same study: intensive discussion was needed to clarify evaluation goals.

The same discipline ought to be asked of interest groups, advisors, and others who draft evaluation language for programs. It is sensible to ask that the questions be specified along with other features of the program.

When Specification Is Not Possible. It will not be possible or desirable to be explicit in every case. To assure that general demands for evaluation are not misinterpreted, the law should provide for a formal assessment of the senses in which the program can be evaluated within one year after the enactment of the legislation.

Regardless of Specification. Regardless of how specific requirements in law are, there is a persistent need for regular dialogue between agency staff and Congressional staff in refining questions and developing agreements on what level of quality of evidence is warranted, and at what cost. The dialogue has occasionally been encouraged in Congressional Committee Reports, by some Congressional staff and by some agency staff. But it is irregular and more heavily dependent on individual preferences than it should be. It is also a demanding process.

Audiences. Because evaluation results may be directed to any number of audiences--the Congress, Department management, interest groups, advisory committees, and so on, there is a clear need for focus. The more audiences there are, the more complex evaluations become.

Pilot Tests of Evaluation Demands. Where there is substantial disagreement about which questions should be addressed and about how the information might be used, pilot evaluations should be undertaken. That is, one mounts formal small scale experiments to determine which of several different evaluation schemes work best. They can be put into the field (a) to determine paperwork burden on respondents, (b) to determine costs of collecting the information, (c) to determine the quality and usefulness of the information, and (d) to clarify language which can be used in statute and regulation.

Use and Authority for Better Evaluation Designs

The authority to use better designs, especially randomized experiments in the interest of relatively unequivocal evaluations of new programs, new program variations, and new program components must be made explicit in law and regulation. For details, see Chapter 5.

By "randomized experiment" here, we mean assigning children, schools, or classrooms randomly to each variation, for instance, and then observing their performance under each regimen. The random assignment is a key feature. It guarantees that, in the long run, comparison of the variations will be fair. This is one of the reasons the design has been used in the Negative Income Tax Experiments, in the Manhattan Bail Bond experiments, in evaluation of T.V. programs such as Sesame Street, the Electric Company, and Free Style, as well as in the evaluation of the effectiveness of medical treatments.

The rationale for the first part of the recommendation, for pilot tests of new programs is that higher quality evaluations are much more feasible before the program is adopted at the national level. Better evaluation designs can be employed, conclusions are less likely to be ambiguous, and political-institutional constraints are less likely to be severe. The introduction of new programs can be staged so that earlier stages constitute pilot tests for the later ones. This may seem terribly mundane to some readers. But recognize that in current political discussion of the proposed Youth Incentives Program, for instance, an enterprise whose probable costs will exceed $850 million per year, there has been no formal attention to pilot testing or staged introduction of the program. Title I compensatory education programs evolved in the same way ten years ago, and we still know pathetically little about effective variations. The simple notion that massive new programs ought to be pilot tested is warranted.

The second part of the recommendation, concerning higher quality evaluation designs, is based on the presumption that we won't learn how to bring about clear detectable changes in the performance of children or schools without more conscientiously designed tests. The justification for the recommendation lies partly in the poor quality of designs used in the field. It is discouragingly easy to find, for example, Congressional Testimony in which a Title I program is declared to be a success by a state legislator because "test scores went up." We do not advocate attempting to estimate program effects in all cases. The process of estimating effects is complicated under the best of conditions. We advocate attention to high quality designs, especially randomized experiments.

At the local level, there are some evaluators with the interest and the skill to employ the design for the sake of fair tests. An obstacle, we believe, is confusion about authority for running such tests. So, for instance, an evaluator offered the opinion that the design is desirable, of course, but in the absence of a clear mandate, could not risk employing it. At the federal level, we believe the authority exists. Indeed, evaluations at that level, such as the one conducted for the Emergency School Assistance Act, have employed state of the art experimental designs. The failure of federal program managers to encourage randomized experiments at the local level is partly because the mandate to do so is not explicit.

Precedent. The usefulness of randomized tests in principle is generally not at issue in discussions about evaluation of new education programs. There is agreement that when experiments are conducted properly, orthodox theory guarantees that long run estimates of program effects will be unbiased. Argument about the uses of the design concerns the idea that randomized experiments are rarely feasible in field settings. Rareness and feasibility are, however, infrequently specified by government policy groups or by individual analysts. Rareness does not establish lack of feasibility and a notable if not large number of field tests have been mounted. Some recent illustrations were covered in Chapter 5. Judging from precedent bald claims that it's impossible to assign individuals or schools or other units randomly to programs for the sake of fair estimates of program effects are unwarranted. It is imperfect evidence in that it doesn't guarantee that an experiment can be mounted successfully in the situation at hand.

Pilot Testing the Experimental Design. We believe that pilot tests of experiments can yield more direct evidence on the feasibility of randomized experiments or other high quality designs. We recommend mounting a small assessment prior to the major field experiments to identify anticipated problems in the field and to resolve them. The main justification for considering such pilot tests is to work out problems beforehand. Randomized experiments fail to be successfully implemented in education as in medicine, economics, etc. because the randomization is incomplete, because the programs are not implemented as advertised, and for other reasons. Pilot tests of the experiment itself can help to avoid unnecessary flaws in implementation.

General Criteria. Lacking dependable precedent and the opportunity for adequate pilot tests of the evaluation design, two general criteria for judging feasibility of randomized experiments seem sensible.

The first criterion turns around the fundamental notion of equity. Where there is an oversupply of eligible recipients for a scarce resource--program services--then randomized assignment of children to the resource seems fair. So, for instance, Vancouver's crisis intervention program for youthful status offenders affords equal opportunity to eligible recipients. Since all could not be accommodated and they are all equally eligible, they are randomly assigned. Experts such as Cook and Campbell argue that randomized experiments are most likely to be carried out successfully when the boon, real or imagined, is in short supply, and the demand for the boon is high. This rationale dovetails neatly with normal managerial constraints. That is, new programs cannot be emplaced all at once and all eligible candidates cannot be served at once. Experiments can then be designed to capitalize on staged introduction of programs or services.

A second criterion concerns settings in which it is politically unacceptable to assign individuals randomly to control conditions despite the fact that we know absolutely nothing about whether a program works relative to no program at all. The ethical, moral, and economic justification for experimenting may be quite irrelevant. In such instances, it is often possible to ameliorate difficulties by comparing program variations against one another, rather than comparing a novel program to an existing one or to no program at all. A "No program" control condition may be an unacceptable political option whether the program fails or not. The most we can reasonably expect then is to choose the invented variation component which works best for the investment.

The idea of testing variations or components rather than testing a program against a control is a compromise. But we believe that getting no information at all on the main one--what are the effects of the program. And the idea is generalizable. In particular, for ongoing programs that have strong public support, it seems sensible to think in terms of randomized assignment to new program variations or randomized of new program components to discover more effective or cheaper versions of the program. This strategy has not been employed by any major ongoing federal education program that we know of. Indeed, it's not common in any social area except the U.S. Census Bureau. In the latter, randomized field tests are periodically run to understand better methods of doing census and surveys.

The most direct action which Congress can take to ameliorate the problem involves any statute which asks that the effects of a new program, new program variation, or new components on children be estimated. We recommend that such statutes include an explicit provision authorizing statistically valid randomized experiments.

For *existing programs*, some explicit authority we believe is necessary to foster *fair* tests. That is, the Secretary should be empowered to: *waive compliance with technical aspects of statutes or regulations for experimental projects which are likely to assist in promoting the statutory objectives*. This would facilitate, for instance, randomized tests of cheaper variations on Title I programs, student loan programs, and the like.

Independent Critique and Secondary Analysis

We recommend, to the Congress and the Department, that major program evaluations be subjected routinely to competent, independent critique and secondary analysis. Mechanisms should be created to permit routine critique of a sample of evaluations produced at the LEA and SEA levels.

By "critique" here, we do not mean adverse commentary. We do mean balanced examination of the quality of the report and judgments about whether recommendations can be sustained by the evidence. Secondary analysis here refers to analysis of raw statistical data, undertaken to improve on the quality of earlier analyses. For details, see Chapter 5.

The origins of this recommendation lie partly in a principle that we should recognize good quality evidence as such, and to properly identify poor evidence. There is also some need to prevent the ingenuous use of poor evidence and to avoid relying unnecessarily on one's confidence in the single evaluation. Furthermore, major evaluations are expensive. It seems sensible to allow the community of policy-makers or their advisors to make the data work repeatedly, at low cost, in secondary analysis. Because evaluations may affect a variety of interest groups, those groups should be given an opportunity to offer competent criticism. Finally, we believe the absence of independent criticism can degrade the importance of good evaluations.

Elements of a System for Critique and Secondary Analysis: National Level. The elements of an effective system for critique and secondary analysis include: (a) explicit institutional policy on rapid disclosure of reports and access to statistical data underlying the reports, (b) a formal mechanism for independent critique or secondary analysis where possible *during* an evaluation, (c) a formal administrative mechanism for independent critique and secondary analysis when evaluation results are submitted, and (d) formal guidelines on reporting and storage of statistical information.

At the national level, elements of policy on reanalysis have already received attention, notably by NIE in supporting research and development on the topic. The GAO has, in its guidelines on impact evaluation, taken the position that access to evaluative data for reanalysis is generally an important consideration. The Department of Health, Education, and Welfare has not had a formal policy on disclosure of statistical data. However, OED has had an unwritten policy and has released data periodically for independent review and secondary analysis. Critiques of particular data sets have been undertaken by the Congressional Budget Office as a part of its efforts to screen studies for quality. These activities are undertaken so as to recognize individual privacy needs. Making policy formal, creating the administrative mechanisms, and testing them are sensible next steps. For details, see Chapter 5.

The problem of rapid access to evaluation reports has been severe. Clearance of OE evaluation reports by the Secretary, according to federal staff members who were interviewed, has been slow at best. We understand that the new Department of Education has adopted the 10 day clearance rule which should improve matters. For details, see Chapter 5.

Informed Criticism. Opinions about the desirability of early independent review of major evaluations and of secondary analysis are not uniform. At least some agency staff reckon that a routine process will generate more heat than light. Assuring competent criticism in this arena is likely to be as difficult as it is in medicine, economics, and other fields. Outcome evaluations are always subject to criticism, especially if the program does not work. Some of that criticism is bound to be specious, dull witted, and self-interested. High quality in design and execution of evaluations offers some protection against unwarranted criticism, but it is unlikely to be sufficient. Adherence to a common set of standards for technical quality also facilitates balanced criticism. But again, this is unlikely to be sufficient. We have not had the time in this investigation to examine prophylactic mechanisms. We believe, however, that openness to criticisms must be given priority, and that some administrative research on reducing mindless criticism should be undertaken.

State and Local Level. A good many local and state "evaluations" provide no more than counts of those served, changes in test scores, and similar information. It is not clear that regular, systematic reanalyses of the raw data underlying these is warranted. It is more clear that samples of reports ought to be critiqued periodically. For details of the content of these reports see Chapter 3. Discussion of their quality appears in Chapter 5

Others, however, are elaborate and these may be used in decisions which are important enough to justify reanalysis. Programs supporting tests of innovative programs fall into this category. Evaluations submitted as part of testimony to the Congress fall into this category.

The main purpose of independent, competent criticism is to assure that the quality of evidence used to inform decisions is recognized. We also expect that this sort of critique will help to improve quality of the exercise, in the long run.

There are a variety of institutional vehicles available to conduct reviews. States with fairly well developed evaluation units are a natural option. States are, for instance, responsible for identifying exemplary Title I programs. Some, like California and Michigan, have review, validation, and dissemination systems for assuring that information about good programs of all kinds is available to LEAs. Such units may not, however, be independent of program offices. Moreover, some field investigation is warranted to determine if evaluation capabilities are sufficient to generate high quality critique. For details on evaluation capabilities see Chapter 4.

The national JDRP is a potential vehicle for critique of samples of evaluations. Its role is now limited to examining evidence volunteered by LEAs and other agencies which believe it is strong enough to sustain frank criticism. And so its mission would have to be expanded. The number of reviewers available on JDRP is not sufficient to review even a small additional sample and so its complement would have to be enlarged.

Technical Assistance Centers supported under Title I constitute another option. But their role is confined to providing advice only when asked about Title I programs. It is not clear that TACs can be regarded as independent reviewers simply because they may provide advice on evaluations in the first instance. Further, not enough is known about their capabilities to warrant expanding their mission without further investigation.

The problems of assuring decent review and reanalysis of evaluation reports is sufficiently important to warrant further examination by federal agency management. That examination should address:

- alternative plans and administrative vehicles for critique and reanalysis,
- alternative sample designs and time frames,
- design of pilot tests for review so as to estimate costs and benefits of a system before it is emplaced, to determine if the effort is indeed justified.

Access to Reports

Effective mechanisms to assure early release of evaluation reports and ready access to reports ought to be created.

The origins of this recommendation lie partly in the idea that evaluation reports offered as a basis for policy, major executive decisions, and oversight should be open to competent criticism and should be accessible to a wide variety of potential users. It stems partly from the difficulty encountered in obtaining reports at the federal level, though this difficulty is far less severe than problems at other levels of government.

Rapid access to reports has over the past few years been impeded by clearance processes within the Education Division. That is, reports issued by a contractor have been reviewed by the Executive Secretariat before release and those reviews have resulted in delays in release without notable improvement in the documents themselves.

The inclusion of a clause in contracts, requiring that permission be sought prior to even discussing an evaluation, is more invidious. It prevents some universities from bidding on evaluations, since the clause runs counter to university standards of intellective independence. It is possible that this proviso reduces the quality of reports by impeding discussion of projects in professional forums. The mechanical difficulties of identifying and obtaining a report or a cluster of reports bearing on a specific evaluation are very tedious. For details see Chapter 5.

Clearance of Reports. The problem of assuring rapid access has been rectified at least in the sense that the Office of the Deputy Assistant Secretary for Evaluation and Management has established a new clearance process. Reports are to be released automatically after 10 days if the Secretary level review has no modifications. The memorandum also permits adjoining criticism to the released document by program managers.

We believe that automatic clearance after a specified period is desirable. We recommend that the practice be maintained regardless of the controversy surrounding a particular report.

The practice of requiring contractors to seek permission for discussing results in public forums has not been examined or resolved. Our recommendation is that no such requirement be imposed in contracts.

Distribution of Information. We suggest the creation of a Department-wide periodical which identifies and abstracts each evaluation report submitted to the Department and submitted by the Secretary to the Congress. We expect this to ameliorate access problems inside and outside the government. At its best, such a periodical will keep the public, the Congress, and staff of the Department abreast of what has been produced and perhaps even why it was produced. Models for this include GAO's Monthly Reports, which summarizes reports issued by the agency.

Responsibility for Distribution. The practice of assigning sole responsibility to the project officer for final reports is not entirely effective. Officers vary in their attention to circulating reports and submitting them to distribution centers such as ERIC. More important, they shift agencies, resign from government service, and otherwise disappear. So do reports, at least at times. Mechanisms must be developed to avoid reliance on the single officer. The options include:

(1) Strengthening internal agency capability for storage of reports.

(2) Assuring that the list of core recipients for reports are included in the reports themselves, or that such a list is publicly available.

(3) Requiring the contractor and the agency to maintain a list of reports, with full citations, generated together with the location of the agency which disseminates it.

(4) Requiring that the recipient of each evaluation executed under contract or grant provide abstracts of reports, reports, or both after 10 day clearance, to ERIC, NTIS, the pertinent education centers and laboratories, CEIS, FEDAC, Congressional staff and support agencies, especially CRS and the GAO

(5) Distribution of each report routinely to every federal evaluation project officer and every evaluation contractor.

Tracking The Use of Evaluations

Our attention to this topic stems partly from the arguments we encountered about whether evaluations are used. The question is not whether they are used: Some are and some are not. The more interesting question is determining how they are used, how often they are used, and how to balance their cost against use. The last question cannot be answered adequately now because the hard answers to the "how" and "how often" questions are fragmentary, and the soft answers are rather too dependent on flawed memory and competing interest. See Chapter 6 for details.

The origins also lie partly in the problem of verifying use or nonuse in special studies such as this one. Turnover of staff responsible for initiating, conducting, and using evaluation is sufficiently high that corroborating use of an evaluation through independent sources is difficult and sometimes impossible. Titles of reports often imply nothing about potential or actual use. The reports are misremembered or forgotten. Incomplete citation is a chronic problem.

The following recommendations are mundane but critical for inexpensive tracking. At best, they will eliminate part of the burden placed on respondents in studies of the use of evaluations.

Better Specification in Reports, Regulations, etc. Failure to specify reports, in both Congressional reports, agency annual reports, regulations, and the like is not prudent. If Congressional Reports, agency annual reports, and the like are to be as useful as possible to the community of thoughtful readers, then references to evaluation should include (a) author, (b) title of the report, (c) date of issuance, and (d) sponsoring agency.

If this is not possible, then merely hiring an inexpensive, bright graduate student to build a specific reference list for each report of the half dozen or so Committees most pertinent to educational evaluation would suffice, so long as access to the list and dissemination of the list was assured.

The recommendation applies to both Congressional Committee Reports and to major agency documents such as OE's Annual Evaluation Report and policy statements. CBO documents are somewhat more conscientious, and GAO documents normally carry at least part of the information suggested. Because we do not have access to CRS documents, we cannot make a judgment. It applies also to proposed and final regulations issued in the Federal Register since evaluations do result in regulation changes but are rarely recognized completely in the prose describing changes. An illustration of exemplary practice for this last is the recent modification of recent regulations on day care.

The practice of recognizing evaluations explicitly when they have been useful in deliberations of Congress and at the executive level is admirable. It ought to be continued for three reasons:

- it identifies what is useful, so guiding the agencies in the long run if not the short,
- it rewards those who perform well,
- it exhibits some integrity to an occasionally cynical audience.

The practice of recognizing good evaluations which are used is not uniform, however. The sponsoring agency is not given credit and so forth on account of time and resource constraints. That problem is serious enough to discourage some staff even if it is not sufficient to demoralize them.

More conscientious attention to recognizing useful evaluations and more conscientious attention to recognizing useless evaluations in committee reports and the like would help.

Improvements of OE's Annual Evaluation Report. The Annual Evaluation Report enumerates uses of evaluations completed by OE, is important, and ought to be the best possible. Simple options for improvement include:

- The report on use should provide specific citation of each evaluation report, its author, title, date of issue and issuing agency. Otherwise, it's impossible for the reader to verify that a report has been issued much less that it has been used.
- The report on use should provide specific citation of hearings or Congressional reports in which an evaluation report is mentioned or used and specific citation of regulations which are said to have been changed on the basis of evaluation results. It should cite regulations which are proposed or created as a result of the evaluation. Otherwise,verifying claims of use is difficult or impossible.
- The contributors to the section on use of evaluations should be acknowledged to permit verification and corroboration.
- The Annual Evaluation Report's perspective on use ought to be reexamined to identify flaws in indicators of use, such as citations of hearings, and the possible biases in them. Ignoring agencies apart from Congress makes it likely that use of evaluation results is understated. Very little information on management uses, apart from regulations, is provided.
- Evaluations for which it is difficult to find verifiable evidence on their use should be identified. Evaluations which are virtually useless two years after production should be identified explicitly.

The issues which ought to be addressed in future examination of uses are: Would reporting other than annually make sense?

Identifying the Recipients of Reports. Major evaluation reports should have appended to them a list of the individuals to whom the report was sent and their affiliations. This will facilitate tracking use simply by making potential users or audiences clear and it will facilitate our understanding of misdirected effort. The practice of appending reader lists to reports is current at the Office of Naval Research. The practice does appear to be feasible for at least major evaluation reports.

Where enumeration of members of the audience is not feasible, then the lists commonly generated internally and used as a basis for sending reports to individuals ought to be accessible. The existence of such lists, their title, and source should be identified in major evaluation reports.

Tracking Management Changes. Very little systematic, publicly available evidence is available on the nature and frequency of managerial uses of evaluation. Moreover, there is no general mechanism for regularly following up on whether problems identified in an evaluation have been rectified. Follow-up does occur episodically, through questions addressed to managers at Committee Hearings for instance. But we have been unable to identify any special, orderly record-keeping on the matter.

We recommend that a simple examination of alternative mechanisms be undertaken to determine if a cheap follow-up system can be developed, and to determine how such mechanisms can be field tested.

Local and State. We have not investigated state uses of evaluations sufficiently to make recommendations on tracking mechanisms at that level. However, two features of some local and state efforts are worth considering by both federal and state agencies. Some states such as Massachusetts and Michigan require that in local reports to the state the uses to which evaluations are put be reported regularly. Those reports are, in principle, a vehicle for tracking use, and occasionally synthesis. We do not know enough about the quality of reporting in this arena. But we believe it ought to be examined. And where some states are found to have developed especially efficient ways to accomplish this task, the procedures ought to be made available to other state and federal agencies. The alternative to regular reporting is a special survery undertaken to obtain periodically a better picture of uses than one could obtain in reports. At least one state, California, has tried this option, and the results are informative.

Standards and Guidelines. Current guidelines can be exploited in designing evaluations and in making crude judgments about quality of an evaluation report. But they are not equally relevant to all types of evaluation, and they are not appropriate for inclusion in law or regulation. They should be recognized in policy statements, internal guidelines, and other flexible directives. See Chapter 5 for details.

Guidelines have been developed to guide design and to facilitate judgments about evaluations. Most focus on planned efforts to assay program effectiveness, not on routine reporting labelled "evaluation."

The guidelines are very general as any set of guidelines on completeness and quality of evidence must be, given the variety of forms which evaluation may take. It is sensible, for instance, to expect that an evaluation which purports to estimate a program's effects on children cover pertinent topics: evaluation design, source and quality of information, competing explanations, and so on. These elements are part of most good guidelines. But they are, of course, no substitute for training and judgment. Moreover, the sensible interpretation of guidelines requires some expertise.

Guidelines have been developed by the U.S. General Accounting Office, the Evaluation Research Society, and the independent Joint Committee on Standards for Educational Program Evaluation. Standards are embodied in manuals used by the USOE-NIE Joint Dissemination Review in assessing educational worth of new programs and the evidence sustaining judgements about worth. There is substantial overlap in topical coverage of all these. Moreover, the topical coverage overlaps with standards used in choosing designs for major national evaluations and grants for evaluative work supported by NIE.

The main justification for recommending that guidelines be recognized is that we believe they can be useful in clarifying what is meant by quality, in informing the public about what can generally be expected of evaluation. Guidelines may also be of some assistance in protecting the competent evaluation from gratuitous criticism, and in identifying the worst cases of inept evaluation. Finally, they can be useful in reviewing proposals made by LEAs for programs which require special evaluation, such as bilingual education.

National Level. We recommend that guidelines be formally recognized as such by agency executives and by Congressional Committee staff. They have already been recognized by evaluation staffers within the education agencies and GAO; indeed, agency and GAO staffers contributed to their development. By recognition here, we mean formal acknowledgement of the existence of guidelines, some effort to assure that pertinent staff know about them, and some effort to test the guidelines in the field. It would not be difficult to incorporate short reviews of guidelines in training programs and seminars on evaluation, run by the CRS, the GAO, or the Federal Executive Institute.

State and Local Level. It is reasonable to assure that SEAs and LEAs know about development of guidelines, to make guidelines available, and to encourage tests of guidelines at the local level. Guidelines can, for instance, be cited in RFPs and grant material without demanding ascription to them. They may be made available through special purpose information clearinghouses, such as the one for bilingual education, and the general purpose ones, such as ERIC.

It is reasonable to encourage their use, not require it, in the interest of fostering better quality evaluations and protecting competence. That encouragement can be given through federal and state agency offices which disburse funds for innovative programs.

Responsibility for advising the public, administrators, school boards, and the like currently rests with evaluation staff at local and state levels. It is not unreasonable to urge that they make guidelines available to these audiences for evaluation results. The guidelines are pertinent, however, to the minority of LEAs which do more than simple monitoring.

Field Tests. We do not recommend incorporating guidelines into law or regulation. Only some aspects of guidelines have been field tested. And regardless of how reasonable they appear to be in principle, their costs and benefits need to be better established before they are generally required. Moreover, it is sensible to determine their susceptibility to incompetent interpretation, misinterpretation, and corruption. Finally, guidelines will change a bit as the state of the art in evaluation develops. And formal tests may help to avoid prematurely rigid posture on what constitutes quality.

Caveats. Contemporary guidelines cannot be simply applied to evaluation reports produced by LEAs in response to federal or state reporting requirements. In the first place, reports differ appreciably in content depending on audience. Reports made to Parent Advisory Committees in Title I programs, for instance, contain information which differs in depth and in kind from information provided to states, for instance. Second, requirements are minimal. Any review of what is produced to fill requirements is not likely to be a useful target for guidelines simply because more elaborate reports may and do exist.

Estimating the Effect of Programs

The general expectation that all local, state, and federal education agencies will produce clear evidence on the effects of programs should be abandoned. The emphasis should be placed on finding better variations on programs in LEAs and SEAs which have the resources to plan and execute fair field tests and on well designed federal tests.

Measuring growth of children in intellective achievement, in personal development, and other areas is often warranted. But the practice of attributing growth to a program on the basis of these data alone is not warranted simply because there are so many competing explanations for growth or any change. Local evaluation designs rarely recognize competing explanations. See Chapter 5 for details.

The demand for information about how much a program affects children must recognize that clearly interpretable estimates depend on evaluation designs which accommodate competing explanations. Those designs are not often feasible in local settings. Technical assistance is no substitute for resources, interest, or those designs. Moreover, estimating effects at the local level often has lower priority than providing services which children and their parents want.

The demand for estimates of effect on children induces a kind of benign hypocrisy among some staffers, administrators, and local contractors responsible for programs and evaluations. An increase in test scores is treated as evidence that the program "works." The conscientious members of each camp will admit that other explanations are possible--normal growth, for instance. But they will also admit, and we agree, that separating out the influence of the program from other influences is not possible without a great deal of managerial, legal, and technical effort and may be impossible despite those efforts. The admission does not appear frequently in evaluation reports on Title I programs, vocational education, and bilingual education.

Judging from our site visits, LEAs and SEAs are interested in testing cheaper varieties of programs, program components, and the like and some of these are capable of doing this well. It is sensible to capitalize on that interest and expertise _if_ the evaluations of these are well designed. To the extent possible, contracts for doing so ought to be made available. Funds are available through Title IV-C and some NIE programs. They can lead to better understanding of what works, what works more inexpensively, and to the dissemination of the products through the JDRP-NDN system. The effort may have to be augmented with assistance from universities, private contractors, technical assistance centers or others. But these are not substitutes for in-house staff and for strong administrative support of fair tests from administrators and oversight groups.

The national interest in understanding effects of _new_ programs, as well as the quality of delivery, needs to be recognized and reiterated. The conduct of pilot tests of new programs should be supported where they are feasible and appropriate.

The origins for this recommendation stem partly from the progress made over the past ten years in mounting field tests of new programs, program variations, and program components. There have been imperfections and failures in these tests to be sure. The execution of good outcome studies is exceedingly difficult. These problems should not be regarded as excuses to avoid the virtue of understanding effects. The public interest in evidence of this kind in education or in other areas such as medicine and economics has not been consistent. Planned tests are always vulnerable on this account as well as on account of their youth.

The questions about how money is spent and to whom services are delivered, so-called process evaluation or implementation studies, are also important. Judging by recent work, the emphasis on this simple information has been understated. There must be some opportunity however to obtain more than body counts, to supply more than nominal statement of where dollars go and who receives the services. The character of services is often poorly understood. Any such investigation will _not_ help one understand whether services produce more notable effects than cheap competitors or no service at all, of course.

At the national level, we believe it is appropriate for public leaders to recognize efforts of this kind at all levels of government. Conscientious tests at the local level are often ignored but do reflect an interest in evidence. That interest should be encouraged. Conscientious tests at the regional or federal level are often not recognized at the local level and there is a need to make these results more available to them. There is some need to assure that properly controlled trials can be run and agencies should be empowered to run them and to suggest provisions in law which will facilitate running them.

Facilitating Integrity

Evaluation often engenders concern among those whose program is evaluated. And this in turn provokes the evaluator concerned if she or he is under the supervision of the program manager. Consequently, maintaining integrity may be difficult. Evaluation does demand some fortitude as well as technical and political expertise.

The following list of options was developed to understand how one might facilitate integrity at federal, state, and local levels. We have capitalized on some experience outside education. We have not had the time to adequately explore each option. But we believe they are worth considering.

Option: Posture at the Policy, Management, and Oversight Levels of Government

There is some argument for the view that administrators of new and innovative projects should not be judged solely on the basis of the success of the program for which they are responsible. Many educational projects are high risk ventures. And their failure is often if not always beyond the control of any individual or institution. It is important to understand why we fail. Program managers and their staffs, then, should be judged on the quality of evidence bearing on a program, regardless, of whether one finds the program itself is a success. To be effective, that view would have to prevail at national, state, and local levels. Many readers will recognize that it also runs against social norms: it is simpler to advocate or oppose than it is to seek balanced information about programs. If this posture is adopted generally, it will take a good deal of time to become routine, perhaps as long as it took for the United States to infuse integrity routinely into U.S. censuses--about 100 years.

Option: Design of Evaluations of New Programs

It is sometimes possible to accommodate fear of evaluation through design of evaluations. One of several simple ways of doing so is not to evaluate program A by comparing it to no program at all. Rather one ought to compare variation A of the program to variation B, where each variation has identical objectives but they differ in cost, approach, or other respects. "No program at all" is often not a politically viable option if A fails. Indeed, it's prudent to compare A against B if one believes in planning: If A and B are equally effective, then one has a contingency plan, choosing B if it's less expensive, and recognizing the high risk of any innovative social program.

The difficulty with the option is that we often lack the imagination or resources to invent B. And, of course, it provides no information on effects of A relative to no program at all.

Option: External Review

One way to assure that incompetent evaluations and competent evaluations are properly labelled as such is to subject completed evaluations to external reviews. The tactic is consistent with the aims of the education agencies, the U.S. General Accounting Office, and other agencies with an interest in quality and standards of evidence. It is consistent with the recent trend toward secondary analysis of program evaluation data, conducted by independent academic institutions. The latter option has been used by, among others, the U.S. Office of Education, the National Institute of Education, the Law Enforcement Assistance Administration, and other agencies in the United States. A variant on the tactic has been tried by individual researchers in Pakistan in reviewing evaluations sponsored by government. But there, as in other developing countries, the matter receives no attention.

This option cannot assure directly that evaluations done with integrity will be rewarded. Gratuitous criticism emerges quickly. It should make it more likely, however, that the poor evaluations are recognized as such and are not rewarded.

Option: Joint Dissemination and Review Panel Approaches

Consider a review board with clearly defined standards for examining the quality of evaluations, and which examine quality in response to a request from the program manager. A main objective of this panel or board is to officially verify that evidence is good and the program, if effective, deserves to be disseminated. Further, such a seal of approval can become a device for obtaining more money for similar projects from an agency. Both official recognition and the opportunity to apply for dissemination funds are appreciated, we believe, by competent evaluators and program developers.

Such a system has been operating with some success by OE and NIE. The joint panel reviews educational products, basing review solely on evidence which conforms to well articulated standards. The approval makes them eligible for specially budgeted money earmarked for expansion, dissemination, and other purposes. The eligible programs compete for additional funds but with less competition than normal and more likelihood of success.

Option: Recognizing Mislabelling and Deception

Bureaucratic detection of mislabelling and deception is generally a high art form. It is not yet as well developed in the evaluation arena as it should be. To take the simplest case, descriptive surveys, needs assessment surveys, studies of management and operations, and the like are labelled "impact evaluations" or implied to be sufficient for impact estimation. The former are easy by comparison to impact studies, they differ in function and use, and they must not be confused. To the extent that they are, people who do a fine job on difficult enterprises may be inadvertently hurt by people doing a good job on easier tasks.

Option: Monitoring and Evaluating Use of Evaluation

One of the major concerns registered by program managers in local and regional education agencies on U.S. Office of Education sponsored programs and elsewhere concerns the uses of evaluation. Evaluations are supposed to be used after all, if they are good. It is in the program manager's interest to understand probable use, and moreover to protect against ingenuous use.

Ingenuous use is possible of course. Many managers, legislative staffers, and so on have little understanding of the quality of evidence and so may naively rely on poor data. That result may affect those with good evidence negatively. Perhaps more important, the quality of utilization will vary depending on the experience of the user. It is simply not easy to capitalize on evidence easily in the interest of making decisions or setting policy. One may, for example, decide on the basis of early evaluation results that a project is ineffective when long term results can show that the program is effective. One may decide that the project in site 1 is ineffective when the project could or would work at site 2. And so on.

Some mechanisms must be invented to encourage, assure, and monitor the high quality use of evaluation results. Several suggest themselves. Training might be useful. This may include case study approaches as in Harvard's MBA programs. The problem can be ameliorated partly through invention of cheap monitoring devices for identifying instances of information use, e.g., periodic telephone calls or cables to agency staff to ask whether information has been used, how it was used and who used it, and verification of the use.

Option: Explicit Policy on Independence

There is no substitute at the national, state, or local levels for policy on relative independence of the evaluator.

That policy may involve bureaucratic independence, notably eliminating clearance requirements for conversation or disclosure of reports to any group. It may involve administrative independence, notably by assuring that the evaluator report to an individual other than the program manager.

It may involve fiscal independence, notably by assuring that funds earmarked for evaluation are channeled through the evaluation unit, by setting salaries for the unit independent of salaries for program operating units, and other methods.

It may involve political independence, for example, through the bipartisan approval of director of evaluation in the same spirit as appointments are approved for Inspector General and Comptroller General.

Evaluation Capabilities

The primary reasons for suggesting that demands for evaluation be preceded by "capabilities assessment", particularly at the state and locals levels, are as follows:

First, identifying who is and who is not an evaluator, much less the appropriate competency level, is often difficult. Depending on the program and the assigned tasks, program staff, evaluation unit staff, outside contractors, or graduate students may have evaluation responsibility. Second, because the field is less than 15 years old, few institutions offer formal certification in the area. There is considerable debate about training, and graudate curricula vary in emphasis across institutions.

More important, the skills and talents required of "evaluators" in LEAs and SEAs differ, depending on evaluation activity. When evaluation involves just meeting federal reporting requirements, the skills demanded do not require advanced graduate training. But some technical common sense is essential. However, when evaluation activities go beyond the minimum reporting requirements, the level and sophistication of required skills multiply quickly. These two types of activities and the capability demands should receive separate consideration in law, regulation, and evaluation policy.

By capabilities assessment here we mean systematic attempt to describe what kinds of skills are required for what kinds of tasks. For national demands for evaluative information, this may involve intensive field research - task analyses - of good performers. But it need not be elaborate. Observing what people do is better, but more expensive of course, than merely asking them how they did it.

Meeting Federal Evaluation Reporting Requirements. It cannot be expected that all state and local education agencies have the capabilities necessary to adequately comply with federal evaluation reporting requirements. Often, program staff in these agencies--individuals with responsibilities other than evaluation--assume responsibility for reporting activities. These persons were not necessarily hired for their evaluation expertise. Consequently, technical assistance in evaluation should be provided so that agencies can adequately fulfill federal evaluation reporting requirements. However, technical assistance can be provided in a variety of ways:

(1) At the minimum, the sponsoring agency should have direct access to evaluation unit staff with explicit responsibility for training in evaluation. These individuals can develop appropriate guidelines for evaluation, arrange evaluation workshops for individuals who must complete these requirements, and select the proper stragegy for providing technical assistance. Federal program agencies without these resources should consider creating specific job positions in evaluation.

(2) Adequate resources can be channeled to SEAs that administer these federal grant programs to permit them to provide easily accessible and expert technical assistance in evaluation.

(3) Federally supported and administered centers such as those existing for Title I evaluation can be established to assist states and local education agencies in meeting federal reporting requirements. One appraoch is to expand services of the present TACs to include provision of evaluation assistance for other federal programs.

Technical assistance in evaluation involves not only instruction and guidance in the actual conduct of evaluation, i.e., selection of program participants, the use of tests, and the completion of federal reporting forms. It also involves assistance in deciding who will evaluate. For example, districts that have capable evaluation units should be encouraged to use the services of the unit for all program evaluation needs. Small districts which do not have the resources to form their own research and evaluation unit may be instructed in other options, e.g., the formation of a consortium to hire competent evaluation staff who serve more than one district. Regional assistance centers can be developed or augmented in order to better provide technical assistance in evaluation. When outside contractors are employed, guidelines must be developed so that program staff and district selection boards can choose the most competent individuals, be sensitive to the types of skills required, and their rights in contractual arrangements.

Going Beyond Federal Evaluation Reporting Requirements. Meeting federal evaluation reporting requirements constitutes minimum acceptable performance for evaluation activities associated with federal programs. Expecting only minimum performance is short-sighted. Some districts and states often attempt to go beyond these requirements. And, if competently executed, these activities can improve the quality of informaton submitted to the federal agencies, Congress, and to such other audiences as Parent Advisory Councils and school boards.

We believe that providing more opportunities to those LEAs and SEAs with interest and capabilities in evaluation is warranted. At the state level, this can be accomplished through such existing mechanisms as the monies targeted for improving state capabilities and state refinement grants for Title I evaluation supported by Section 183(c). These funding mechanisms should be supported. Dissemination of demonstrated improvements in evaluation practices developed by SEAs through such contracts should be promoted.

The improvement of local education agency capabilities deserves more attention than it has received in terms of discretionary evaluation activities. While some of this can be accomplished through an expanded SEA role, other methods can be more specifically targeted at LEAs and supported directly from the federal government. Options include:

(1) Expanding the program of direct contracts to LEAs for evaluation-related activities can be instituted. These grants should allow LEAs to apply for and receive funds to engage in additional evaluation activities for federal programs or research on ways to improve evaluation methods.

(2) Making available grants to LEAs/SEAs to foster university LEA relationships can be developed. This might include funding for training programs jointly sponsored by academic institutions and LEAs/SEAs. This would not only provide training for agency personnel but also improve the quality of evaluation programs in universities by allowing students to participate in actual evaluations. In addition, university conducted workshops could be supported as an avenue of continuing education for education agency personnel. There may also be an opportunity to award matching monies for SEA/LEA investment in such arrangements as "an endowed chair" whereby university faculty can spend a period of time in these agencies conducting evaluations and designing procedures which will remain after their departure.

8. REFERENCES

Aaron, H. Statement. In: Hearings on costs, management, and utilization of human resources program evaluation, 1977. U.S. Senate Committee on Human Resources, 95th Congress, 1st Session, October 6 and 27, 1977. Washington, D.C.: U.S. Government Printing Office, 1978.

Abt Associates, Inc. National day care study: Preliminary findings and their implications. Cambridge, MA: Abt, 1978.

Abt Associates, Inc. Study of the Emergency School Aid Act Magnet School Program. Cambridge, MA: Abt, 1979.

Abt, C.C. Government constraints on evaluation quality. In: L.E. Datta and R. Perloff (Eds.) Improving evaluations. Beverly Hills, CA: Sage, 1979, pp. 43-52.

Abt, C.C. Reversing the productivity decline with social sciences R&D. Paper presented to the New England Venture Capital Association, 1980 (Avaliable from the author, Abt Associates, Cambridge, MA).

Abt, C.C., & Abt, W.P. Review of Academic and Entrepreneurial Research (by I.N. Bernstein & H.E. Freeman). Social Science and Medicine, 1976, 10, 344-345.

Abt, C.C. What the future holds for children in the T.V. Computer age: Unprecedented promises and intolerable threats to child development. Keynote address to the National Council for Children and Television Sumposium on Children, Families, and New Video/ Computer Technologies, Princeton, New Jersey, March 10, 1980. (Available from the author, Abt Associates, Cambridge, MA).

Abt, C.C. What's wrong with social science R&D? Leadership failures and opportunities in social research. Unpublished manuscript. Abt Associates, Cambridge, MA, 1980.

Acland, H. Secondary analysis of the Emergency School Assistance Program, Santa Monica: Rand, 1975.

Alkin, M.C., Daillak, R., and White, P. Using evaluations: Does evaluation make a difference? Beverly Hills: Sage Publications, 1979.

Alkin, M.C., Kosecoff, J., Fitz-Gibbon, C. and Seligman, R. Educational decision making: The Title VII experience. CSE Monograph Series #4, 1974.

American Institute for Research, Evaluation of the impact of ESEA Title VII, Spanish/English Education Program. Palo Alto, CA: AIR, 1978.

Anderson, R.B., St. Pierre, R.G., Proper, E.C., and Stebbins, L.B. Pardon us, but what was the question again? A response to the critique of the Follow Through evaluation. Harvard Educational Review, 1978, 48, 161-170.

Anderson, S.B., & Ball, S. *The profession and practice of program evaluation*. San Francisco: Jossey-Bass, 1978.

Anderson, S.B., & Ball, S. Training evaluators and evaluating their competencies. In L. Sechrest and others (Eds.) *Evaluation Studies Review Annual*, 1979, 4, 349-373.

Applied Management Sciences. *Stage I final report for the study of program management in the Campus-based and Basic Grant Programs*. Silver Spring, Maryland: Applied Management Sciences, Inc., 1978.

Applied Management Sciences, Inc. *Study of program management procedure in the basic grants programs.* Silver Spring, MD: Applied Management Sciences, Inc., 1978.

Applied Management Sciences. *A survey of viewership of television series sponsored by ESAA legislation*. Silver Spring, Maryland: Applied Management Sciences, Inc., 1978.

Assistant Secretary for Planning and Evaluation. *Final Report on Follow Through*. Washington, D.C.: Department of Health, Education and Welfare, Office of the Assistant Secretary for Planning and Evaluation, Office of Evaluation and Technical Analysis, 1979, EDC 1012.

Assistant Secretary for Planning and Evaluation. *Impact Aid two years later: An assessment of the program as modified by the 1974 Education Amendments*. Washington, D.C.: U.S. Government Printing Office, 1977.

Assistant Secretary for Planning and Evaluation. *Report on evaluation utilization in the Department of Health, Education, and Welfare*, Washington, D.C.: DHEW, 1980.

Austin, G.R., Rogers, B.G., & Walbesser, H.H. The effectiveness of summer compensatory education: A review of the research. *Review of Educational Research,* 1972, *42*(2), 171-182.

Averch, H.A. et al. *How effective is schooling*? Englewood Cliffs, N.J.: Educational Technology Publications, 1974.

Ball, S., & Bogats, G.A. *Reading with television: An evaluation of the Electric Company*. Princeton, N.J.: Educational Testing Service, 1973.

Ball, S., & Bogatz, G.A. *The first of Sesame Street: An evaluation.* Princeton, NJ: Educational Testing Service, 1970.

Barkdoll, G.L. Type III evaluations: Consultation and consesus. *Public Administration Review*, 1980, March/April, 174-179.

Barnes, R.E. and Ginsburg, A.L. Relevance of the RMC models for Title I policy concerns. *Educational Evaluation and Policy Analysis*, 1979, *1*(2), 7-14.

Barnow, B.S. and Cain, G.G. A reanalysis of the effect of Head Start on cognitive development: Methodological and empirical findings. *Journal of Human Resources*, 1977, *12*, 177-197.

Berge, I.P. Utilization of educational program evaluation for federal policy making: The case of bilingual education. Paper presented at the Evaluation Research Society Convention, Minneapolis, MN, October, 1979.

Berliner, D.C., & Tikunoff, W.J. The California Beginning Teacher Evaluation Study: Overview of the ethnographic study. *Journal of Teacher Education*, 1976, *27*(1), 24-30.

Berman, P. Prepared testimony before the House Subcommittee on Elementary and Secondary Education, for the *Education Amendments of 1978, Hearings, Part 8, ESEA Consolidated Programs*.

Berman, P., and McLaughlin, M.W. *Federal Programs Supporting Educational Change*. Santa Monica: Rand Corporation, 1977 (8 Volumes).

Bernstein, I.N. and Freeman, H.E. *Academic and entrepreneurial research.* New York, NY: Sage, 1975.

Berryman, S.E. and Glennan, T.K. An improved strategy for evaluating federal programs in education. *Rand Note*. Santa Monica: Rand Corporation, August 1979. (NIE-C-300-76-0065).

Bissell, J. Use of educational evaluation and policy studies by the California Legislature. *Education Evaluation and Policy Analysis*, 1979, *3*(1), 29-37.

Bogatz, G.A. and Ball, S. *The second year of Sesame Street: A continuous evaluation*. Volumes 1 and 2. Princeton, New Jersey: Educational Testing Service, 1971.

Boruch, R.F. On common contentions about randomized experiments. In G.V. Glass (Ed.) *Evaluation Studies Review Annual*, 1976, *3*, 158-194.

Boruch, R.F., McSweeny, A.J., & Soderstrom, J. Bibliography: Illustrative randomized field experiments. *Evaluation Quarterly*, 1978, 655-695.

Boruch, R.F., and Wortman, P.M. Implications of educational evaluation for evaluation policy. In: D. Berliner (Ed.) *Review of Research in Education*, 1979, in press.

Boruch, R.F., & Riecken, R.F. Experimental testing of public policy. Boulder, Colorado: Westview, 1975.

Boruch, R.F. and Cecil, J.S. Ensuring the confidentiality of data in social research. Philadelphia, PA: University of Pennsylvania Press, 1979.

Brager, G.L. and Mazza, P. The level of analysis and the level of presentation are not the same. Educational Evaluation and Policy Analysis, 1979, 1(3), 105-106.

Brandl, J. Evaluation and politics. Evaluation, 1978, 6, 6-7.

Braskamp, L.A. and Brown, R.D. (Eds.) Utilization of evaluative research. New Directions for Program Evaluation, 1980, Number 5.

Breger, M. Legal aspects of social research. Unpublished paper, presented at the Symposium on Ethical Issues in Social Research, University of Minnesota, April 1976.

Breglio, V.J., Hinkley, R.H., Beal, R.S. Students' economic and educational status and selection for compensatory education. Technical Report No. 2 of the Sustaining Effects Study, Santa Monica, CA: Decima Research, under contract with System Development Corporation, April 1978.

Brickell, H.M. Data for decisions: An analysis of evaluation data needed by decision makers in educational programs. New York, NY: Policy Studies in Education, 1974.

Brickell, H.M. and Paul, R.H. Promoting and achieving sex equity in vocational education: Assessing the efforts of the National Center for Vocational Education. Columbus, Ohio: NCRVE, 1979.

Brown, B. (Ed.) Found: Long term gains from early intervention. Boulder, Colorado: Westview, 1978.

Budget gives bilingual education $174 million for English emphasis, Education Daily, January 23, 1979.

Bunker, J.P., Barnes, B.A., and Mosteller, F. (Eds.) Costs, risks, and benifits of surgery. New York, NY: Oxford University Press, 1977.

Burke, F.E. Testing and funding: The New Jersey experience. New Directions for Testing and Measurement, 1979, 1, 45-52.

Burkheimer, G.J., Riccobono, J.A., and Weisenbaker, J.M. Evaluation study of the Upward Bound program: A second follow-up. Chapel Hill, NC: Research Triangle Institute, November 1979.

Califano, J.A. Memorandum for Assistant Secretaries, Heads of POCs, Heads of staff offices, Principal Regional Officials, on transmittal of reports to Congress. Office of the Secretary, Department of Health, Education and Welfare, Washington, D.C., April 10, 1978.

Califano, J.A. *Prepared testimony before the House Committee on Education and Labor*. In Hearings on reauthorization of ESEA, Part 3, 1977.

Campbell, D.T., & Erlebacher, A. How regression artifacts in quasiexperimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), *Compensatory education: A national debate.* New York, NY: Brunner/Mazel, 1970.

Campeau, P.L. et al. *Final report: Evaluation of Project Information Package Dissemination and Implementation*. Palo Alto, CA: American Institutes for Research, 1979.

Campeau, P.L., Roberts, A.O.H., Bowers, J.E., Austin, M., and Roberts, S.J.. *The identification and description of exemplary bilingual education programs.* American Institutes for Research, Palo Alto, CA: August 1975. Contract number OEC-0-74-9331 from USOE, DHEW.

Caplan, N. A minimal set of conditions necessary for the utilization of social knowledge in policy formulation at the national level. In C.H. Weiss (Ed.) *Using social research in public policy making*. Lexington, MA: D.C. Heath, 1977, pp. 183-197 (b).

Caplan, N., et al. *The use of social science knowledge in policy decisions at the national level: A report to respondents*. Ann Arbor, Michigan: University of Michigan, Institute for Social Research, 1975.

Carrallo, S.B. The evolution of USOE evaluations of the "TRIO" programs. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA, April 9, 1980.

Carter, L.F. Federal clearance of educational evaluation instruments: Procedural problems and proposed remedies. *Educational Researcher*, 1977, *6*, 7-12.

Carter, L.F. The Sustaining Effects Study: Its purpose, general design and data colleciton. Paper presented at the Annual Meeting of the American Education Research Association, Boston, MA, April 7-11, 1980. (Available from the author, System Development Corporation, Santa Monica, CA).

Caulley, D.N., & Smith, N.L. Program evaluation in state departments of education. *Evaluation News*, 1980, *14*, 37-38.

Chalmers, T.C., Block, J.B., and Lee, S. Controlled studies in clinical cancer research. *New England Journal of Medicine*, 1972, *287*, 75-78.

Chelimsky, E. Differing perspectives of evaluation. New Directions for Program Evaluation, 1978, Summer (2), 1-18.

Chelimsky, E. Evaluating broad aim, large scale social demonstration programs. In F.M. Zweig (Ed.) Evaluation in legislation. Beverly Hills, CA: Sage, 1979, pp. 121-145.

Coelen, C., Grantz, F., and Colore, D. Day care centers in the U.S.: 1976-77. Cambridge, MA: Abt, 1979.

Coleman, J.S., Campbell, E.Q., Hobsen, C.J., McPartland, J., et al. Equality of educational opportunity. Washington, D.C.: U.S. Government Printing Office, 1966.

Coleman, J., Bartot, V., Lewin-Epstein, N., Olson, L. Policy issues and research design: Report to the National Center for Educational Statistics (Contract No. 300-78-0208) by the National Opinion Research Center, Chicago, IL: NORC, October, 1979.

Coleman, J.S. Response to Pettigrew and Green. Harvard Educational Review, 1976, 46, 217-224.

Congressional Budget Office. Childcare and preschool: Options for federal support. Washington, D.C.: CBO, September, 1979.

Congressional Budget Office. Federal student assistance: Issues and options. Washington, D.C.: CBO, March 1980.

Congressional Budget Office. The federal effort for high schools: A preliminary analysis. Staff Draft Analysis. Washington, D.C.: CBO, January 1980.

Conner, R.F., Clay, T., & Hill, P. Evaluation training: Results from the Evaluation Research Society survey. Paper presented at the meeting of the Evaluation Research Society, Minneapolis, MN, October 1979.

Cook, T.D., et al. Sesame Street revisited. New York, NY: Sage, 1975.

Coulson, J.E. National evaluation of the Emergency School Aid Act (ESAA): A review of methodological issues. Journal of Educational Statistics, 1978, 3(1), 1-60.

Crain, R.L. and Mahard, R.E. Desegregation and black achievement. Law and contemporary problems, in press.

Crain, R.L. and others. Southern schools: An evaluation of the Emergency School Assistance Program and of school desegregation. 2 Volumes. Chicago, IL: National Opinion Research Center, 1973.

Crandall, D.P. and others. *Program validation: Four case studies/A brief report on four projects and their experiences before the Joint Dissemination Review Panel.* Andover, Maine (290 S. Main Street): The NETWORK, 1979.

Crocker, S. and others. *Title IV of the Civil Rights Act of 1964: A review of program operations.* Santa Monica, CA: Rand Corporation, 1976.

Cronbach, L.J. and others. *Designing educational evaluations*. Stanford, CA: Stanford Evaluation Consortium, 1978.

Cronbach, L.J., Suppes, P. (Eds.), Coleman, J.S., et al. *Research for tomorrow's schools: Disciplined inquiry for education*. New York, NY: MacMillan, 1969.

Cronbach, L.J. Research on classrooms and schools: Formulation of questions, design and analysis (Stanford Evaluation Consortium Occasional Paper). Unpublished paper, Stanford University, 1976.

Cross, C.T. Title I evaluation - A case study in Congressional frustration. *Educational Evaluation and Policy Analysis*, 1979, *1*, 15-22.

Crowe, M.R., Rice, J., & Walker, J.P. *Evaluation of the Executive High School Internships Program, Final Report*, Center for Vocational Education, Ohio State University, Columbus, Ohio, December 1977.

Daniels, R.S. and O'Neil, C. The reverse RFP. *Educational Evaluation and Policy Analysis*, 1979, *1*(1), 43-44.

Danoff, M., and others. *Evaluation of the impact of ESEA Title VII Spanish/English bilingual education program - overview of study and findings.* Palo Alto, CA: American Institutes for Research, 1978.

Datta, L. "Damn the experts and full speed ahead: An examination of the study of *Federal Programs Supporting Educational Change* as evidence against directed development and for local problem solving." Address presented at the Annual Meeting of the Evaluation Research Society, Minneapolis, MN, 1979. (Available from the author, NIE).

Datta, L. Design of the Head Start Planned variation experiment. In: A.M. Rivlin and P.M. Timpane (Eds.) *Planned variations in education*.. Washington, D.C.: Brookings, 1975, pp. 79-87.

Datta, L. Does it work when it has been tried? And half full or half empty? *Journal of Career Education,* 1976, *2*(3), 38-55.

Datta, L. The external implications of an internal review of effectiveness: The DHEW Education Division's JDRP. In *the Education Division's JDRP: Three Papers.* New York, NY: EPIE Institute, 1977. (ERIC #ED156217).

Datta, L. Front end analysis: Pegasus or shank's mare? *New Directions for Program Evaluation*, 1978, Spring (1), 13-30.

Datta, L. The impact of the Westinghouse/Ohio evaluation on the development of Project Head Start. In C.C. Abt (Ed.) *The evaluation of social programs*. Beverly Hills, CA: Sage, 1976, pp. 129-191 (b).

Datta, L. Interpreting data: A case study from the Career Intern Program Evaluation Presented at Symposium on Evaluation Social Programs and their Dissemination. Michigan State University, East Lansing, MI, May 25-26, 1977.

Datta, L. O thou that bringest the tidings to lions: Reporting findings of educational evaluations. Presented at the Second Annual Johns Hopkins University Symposium on Educational Research, Evaluation Methodology: The State of the Art, Washington, D.C., November 2, 1979 (Available from the Author, Executive Director, Office of Research and Evaluation, School District of Philadelphia, Philadelphia, PA).

David, J.L. *Local uses of Title I evaluations: Executive summary*. Palo Alto, CA: SRI International, July 1978 (Contract HEW-100-77-0095).

David, J.L., and Pelavin, S.H. *Research on the effectiveness of compensatory education programs: A reanalysis of data. Final Report*. (SRI Project URU-4425, Prepared for the National Institute of Education). Menlo Park, CA: SRI International, September 1977.

Davidson, J.L. The research and evaluation unit: Helping your school board make decisions. Paper presented at the Annual Meeting of the American Association of School Administrators, Atlantic City, New Jersey, 1976. (Available from the author, Superintendent, Austin Independent School District, Austin, Texas).

Decima Research. *Study of the sustaining effects of compensatory education on basic skills*. Santa Monica, CA, January 1978.

Doherty, V.W. Living in two worlds. *Educational Evaluation and Policy Analysis*, 1979, *1*(2), 57-65.

Dollar, C.M. and Ambacher, B.I. The national archives and secondary analysis. *New Directions in Program Evaluation*, 1978, *4*, 1-6.

Educational Programs That Work. San Francisco, CA: Far West Laboratory, 1979. (ERIC #ED149441).

Ellis, J. Using measurement in educational decision making. *New Directions for Testing and Measurement*, 1979, *1*, 89-96.

Elmore, R.F. Design of the Follow Through experiment. In: A.M. Rivlin and P.M. Timpane (Eds.) Planned variations in education. Washington, D.C.: Brookings, 1975, pp. 23-45.

Emrick, J.A., Peterson, S.M., and Agarwala-Rogers, R. Evaluation of the National Diffusion Network: Final Report. (SRI Project 4385, Contract 300750360 (OPBE)). Menlo Park, CA: SRI International, May 1977.

Evaluation Research Society. Standards for program evaluation. Unpublished manuscript, November 1979 (available from Scarvia Anderson, Chairperson, ERS Committee on Standards, Educational Testing Service, Atlanta, Georgia).

Evans, J.W. Evaluating educational programs - Are we getting anywhere? Educational Researcher, 1974, 3, 7-12.

Evans, J.W. Motives for experimentation. In R.F. Boruch and H.W. Riecken (Eds.) Experimental testing of public policy. Boulder, CO: Westview, 1975, pp. 15-22.

Evans, J.W. Statement. In U.S. Senate Committee on Human Resources. Oversight hearings on the scope and extent of the federal government's efforts to evaluate its programs. 95th Congress, 1st Session. Washington, D.C.: U.S. Government Printing Office, 1977.

Fairweather, G.W., Sanders, D.H., & Tornatsky, C.G. Creating change in mental health organizations. New York, NY: Pergamon, 1974.

Federal Judicial Center, Research Division. Project Status report. Washington, D.C.: Federal Judicial Center, March 21, 1980.

Federal Register, April 11, 1980. Department of Health, Education and Welfare, Office of Education. Emergency School Aid.

Federal Register, May 16, 1980. Department of Education, Emergency School Aid. Washington, D.C.: U.S. Government Printing Office, 1980.

Federal Register, Volume 45, No. 55, Part V, March 19, 1980, pp. 17870-17885.

Federal Register, April 21, 1980. Department of Health, Education and Welfare, Part V, Office of Education. National Diffusion Network Program: Final Regulations Governing the Award of Grants. Washington, D.C.: U.S. Government Printing Office, 1980.

Federal Register, Volume 44, No. 117, Part III, June 15, 1979, pp. 34754-34781.

Federal Statistical System Project. Issues and options. Unpublished Draft Report. President Reorganization Project for the Federal Statistical System, Washington, D.C., November 30, 1978.

Feinberg, L. "Basic" teaching methods more effective, study says. Washington Post, June 20, 1977, p. 1.

Felix, J.L. Research and Evaluation to improve instruction: The Cincinnati strategy. Educational Evaluation and Policy Analysis, 1979, 1, 57-62.

Feller, I., et al. Scientific and technological information in state legislatures. American Behavioral Scientist, 1979, 22(3), 417-436.

Flaherty, D.L. Final report of the Bellagio conference on privacy, confidentiality, and the use of government microdata for research and statistical purposes. Federal Statistical Reporter, 1978, 8, 274-278.

Florio, D.H. Educational inquiry and the Congress: Unmet potential or undervalued resource? Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA, April 1980.

Floric, D., et al. What do policy makers think of educational research and evaluation? Or do they? Unpublished manuscript. 1979.

Forehand, G.A. and Ragosta, M. A handbook for integrated schooling. Princeton, NJ: Educational Testing Service, 1976.

Forehand, G.A., Ragosta, M. and Rock, D.A. Conditions and processes of effective school desegregation. Princeton, NJ: Educational Testing Service, 1976.

Foskett, W.H. and Fox, H.W. Program evaluation: A manual for legislators and legislative staffs. In U.S. Senate Subcommittee on Oversight Features, Committee on Government Operations. Legislative oversight and program evaluation. (94-2) Washington, D.C.: U.S. Government Printing Office, May 1976, pp. 464-498.

Fox, H.W. Evaluation research and Congressional oversight: A case study of a general phenomenon. Presented at the Annual meeting of the American Association for the Advancement of Science, Denver, CO. February 20-25, 1977. (Available from the author, Industrial College, Washington, D.C.).

Frankel, J., Sharp, L.M., & Biderman, A.D. Performers of research and research-related activities in the field of education. Washington, D.C., Bureau of Social Science Research, 1979.

Gage, N.L. A factorially designed experiment on teacher tutoring, structuring, soliciting, and reacting. Journal of Teacher Education, 1975, 27(1), 35-38.

Gerard, H.B., & Miller, N. *School desegregation*. New York, NY: Plenum, 1975.

Gilbert, G.R., and Conklin, P.J. *Evaluation management: A selection of readings*. Charlottesville, Virginia: Federal Executive Institute, U.S. Office of Personnel Management, 1979 (FEI B-12).

Gilbert, J.P., Light, R.J., and Mosteller, F. Assessing social innovations: An empirical base for policy. In C.A. Bennett & A.A. Lumsdaine (Eds.) *Evaluation and experiment*. New York; Academic Press, 1975.

Gilbert, J.P., McPeek, B., & Mosteller, F. Progress in surgery and anesthesia: Benefits and risks of innovative therapy. In: J.P. Bunker, B.A. Barnes, and F. Mosteller (Eds.) *Costs, risks, and benefits of surgery*. New York, NY: Oxford University Press, 1977, pp. 124-169.

Gilbert, J.P., Mosteller, F., & Tukey, J. Steady social progress requires quantitative evaluation to be searching. In C.C. Abt (Ed.) *The evaluation of social programs.* Beverly Hills, CA: Sage, 1976, pp. 295-312.

Gilette, B. Letter to members of the Iowa State Sex Equity Advisory Committee, April 15, 1980.

Glass, G.V. Primary, secondary, and meta-analysis of research. *Educational Researcher*, 1976, *5*(10), 3-8.

Glass, G.V. Policy for the unpredictable (uncertainty research and policy). *Educational Researcher*, 1979, *8*(9), 12-14.

Glass, G.V. and Ellett, F.S. Evaluation research. *Annual Review of Psychology*, 1980, *31*, 211-228.

Glass, G.V. When educators set standards. In E.L. Baker and E.S. Quellmalz (Eds.) *Educational testing and evaluation: Design, analysis, and policy*. Beverly Hills, CA: Sage Publications, 1980.

Glennan, T.K. Review of "Educational policy and international assessment: Implications of the IEA surveys of achievement." *Harvard Educational Review*, 1976, *46*, 118-122.

Goor, J. *School district needs for Title I evaluation technical assistance*. (NCES 80-231). Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1979.

Gordon, A.C. and Heinz, J.P. (Eds.) *Public access to information*. New Brunswick, NJ: Transaction, 1979.

Gordon, G. and Morse, E.V. Evaluation research. *Annual Review of Sociology*, 1975, *1*, 339-361.

Gramlich, E.M. and Koshel, P.P. Educational performance contracting, Washington, D.C.: Brookings, 1975.

Guba, E.G., and Stufflebeam, D.L. Evaluation: The process of stimulating, aiding, and abetting insightful action. Paper presented at the Second Phi Delta Kappa National Symposium for Professors of Educational Research, Boulder, CO, November.

Hall, G.E. and Loucks, S.F. A developmental model for determining whether the treatment is actually implemented. American Educational Research Journal, 1977, 14(3), 263-276.

Hamilton, J.A. and Mitchell, A.M. Final technical report: Identification of evaluated, exemplary activities in career education (K-12). Palo Alto, CA: American Institutes for Research, 1978.

Harrison, L.R., and Dahl, P.R. Executive summary: Vocational education equity study. Palo Alto, CA: American Institutes for Research, 1979.

Hart, G. The California Pupil Proficiency Law as viewed by its author. Phi Delta Kappan, May 1978.

Havens, H.S. with Marvin, K.E., Lauve, R.F, and Varholy, S.J. Statement on expenditures, problems, and prospects of management and utilization of program evaluation. Program Analysis Division, U.S. GAO, Washington, D.C., October 6, 1977.

Hawkridge, D.G., Chalupsky, A.B., & Roberts, A.D.H. A study of selected exemplary programs for the education of disadvantaged children. Parts I and II. Palo Alto, CA: American Institutes for Research, 1968.

Hedrick, T.E., Boruch, R.F., and Ross, J. On ensuring availability of evaluative data for secondary analysis. Policy Sciences, 1978, 9, 259-280.

Hendrickson, L. and Stephens, C. Anatomy of a major evaluation: Context, procedures, and results of a two year evaluation of a school district's reading instruction. Educational Evaluation and Policy Analysis, 1979, 1(3), 77-84.

Hill, P. Evaluating education programs for federal policy makers: Lessons from the NIE Compensatory Education Study. In J. Pincus (Ed.) Educational evaluation in the public policy setting. Santa Monica, CA: Rand Corporation, May 1980. (R-2502-RC).

Hoepfner, R. Summer "drop-off," Summer School & its effectiveness. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA, April 7-11, 1980 (Available from the author, System Development Corporation - SES Study Group, Santa Monica, CA).

Hoepfner, R., Wellich, J., and Zagorski, H. Report #1. The sample for sustaining effects study and projections of its characteristics to the national population. Santa Monica, CA, March, 1977.

Holley, F.M. The role, importance, and problems of local/state evaluation in educational development. Unpublished report, 1980. (Available from the author, Office of Research and Evaluation, Austin Independent School District, Austin, Texas).

Holley, F.M., Davidson, J.L., Lee, A.M., and Matuszek, P. Models for the delivery of school district evaluation: Service or accountability. Paper presented at the annual meeting of the American Educational Research Association, April 1976 (Available from the authors, Office of Research and Evaluation, Austin Independent School District, Austin, Texas).

Holley, F.M. and Ligon, G. Time: Today it translates to dollars. Educational Evaluation and Policy Analysis, 1979, 1(3), 109-112.

Hope Associates, Inc. Performance review of USOE's Title I evaluation Technical Assistance program. Washington, D.C.: Hope Associates, 1979. (DHEW Contract 300-78-0355. Reviewers: Jason Millman, William Paisley, W.T. Rogers, James R. Sanders, Frank B. Womer, Keith Baker, Project Officer).

Horst, D.P., Tallmadge, G.K., & Wood, C.T. A practical guide to measuring project impact on student achievement. Washington, D.C.: U.S. DHEW, 1975.

House, E.R. The objectivity, fairness, and justice of federal evaluation policy as reflected in the Follow Through evaluation. Educational Evaluation and Policy Analysis, 1979, 1(1), 28-42.

House, E.R., Glass, G.V., McLean, L.D., and Walker, D.F. No simple answer: Critique of the Follow Through evaluation. Harvard Educational Review, 1978, 48, 128-160.

Ingle, M., & Klauss, R. Classifying evaluator competencies: Issues, approaches and applications. Symposium paper presented at the meeting of the Evaluation Research Society, Minneapolis, MN, October 1979.

Jackson, G.B. The research evidence on the effects of grade retention. Review of Educational Research, 1975, 45(1), 613-635.

Jacobs, J.N. and Felix, J.L. Improvement despite austerity: R&D in Cincinnati. Educational Evaluation and Policy Analysis, 1979, 1(2), 47-52.

Jamison, D., Suppes, P., & Wells, S. The effectiveness of alternative instructional media: A survey. Review of Educational Research, 1974, 44(1), 1-67.

Joint Committee on Standards for Educational Evaluation. Draft Standards for evaluations of educational programs, products, and materials. Kalamazoo, MI: Western Michigan University, 1978.

Kean, M.H. Comments on "O thou that bringest the tidings to lions: Reporting the findings of educational evaluations." Presented at the Second Annual Johns Hopkins University National Symposium on Educational Research, Evaluation Methodology. The State of the Art, Washington, D.C., November 2, 1979 (Available from the author, Executive Director, Office of Research and Evaluation, School District of Philadelphia, Philadelphia, PA).

Kennedy, M.M. Developing an evaluation plan for Public Law 94-142. New Directions for Program Evaluation, 1978, 2, 19-38.

Kennedy, M.M. Designing a flexible evaluation: The evaluation of Public Law 94-142. Presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, April 1979.

Kilpatrick, J.J. Basics better in education, Cambridge group finds. Boston Evening Globe, July 1, 1977, p. 19.

Kloman, E.H. (Ed.) Cases in accountability: The work of the GAO, Boulder, CO: Westview Press, 1979.

Knezo, G. Program evaluation concepts. In: Hearings on legislative oversight and program evaluation. Subcommittee on Oversight Procedures, Committee on Government Operations, U.S. Senate, 94th Congress, 2nd Session, Washington, D.C.: U.S. Government Printing Office, May 1976, pp. 55-62.

Kruskal, W. Formulas, numbers, words: Statistics in prose. American Scholar, 1978, 47(2), 223-229.

Kruskal, W.B., & Mosteller, F. Representative sampling. Technical Report No. 47. Chicago, IL: Statistics Department, University of Chicago, (September) 1977.

La Tour, K. More high marks for the 3 r's: Basic skills found lagging in "open" classrooms. National Observer, July 4, 1977, p. 6.

Lee, D. Understanding compensatory education through in-depth study. Paper presented at the American Educational Research Association Annual Meeting, Boston, April 7-11, 1980 (Available from the author, System Development Corporation, Sustaining Effects Study Group, Santa Monica, CA).

Lesnick, H. and Langsdorf, M. The call for policy-relevant evaluation: A federal agency responds. New Directions for Program Evaluations, 1978, Summer (2), 39-54.

Light, R.J. Capitalizing on variation: How conflicting research findings can be helpful for policy. Educational Researcher, 1979, 8(9), 7-11.

Linn, R.L. Validity of inferences based on the proposed Title I evaluation models. Educational Evaluation and Policy Analysis, 1979, 1(2), 23-32.

Lora vs. the Board of Education of the City of New York, 74 F.R.D. 565 (E.D.N.Y., 1977).

Lukas, C.V. Problems in implementing Head Start Planned variation models. In: A.M. Rivlin and P.M. Timpane (Eds.) Planned variations in education. Washington, D.C.: Brookings, 1975.

Lyons, C., Doscher, L., McGranahan,P. and Williams, R. Evaluation and School districts. (Preliminary Report). Los Angeles, CA: Center for the Study of Evaluation, December 1978.

Mathis, W. A national survey of state educational agency research organizations. Paper presented at the meeting of the American Educational Research Association, San Francisco, April 1979.

Manning, C. and others. District review of vocational education: Follow-up three studies. Napa County, CA: Vocational Field Operations Section, Office of the Superintendent of Schools, September 1975.

Manning, C. District review of vocational education: Follow-up study of 19 school districts and four regional occupational programs reviewed during 1974-75. Napa County, CA: Vocational Field Operations Section, Office of the Superintendent of Schools, 1976.

Marvin, K.E. Evaluation for Congressional committees. In F.M. Zweig (Ed.) Evaluation in legislation. Beverly Hills, CA: Sage, 1979, pp. 45-56.

Matuszek, P., and Lee, A.M. Evaluating evaluation: Paper presented at the Annual Meeting of the American Educational Research Association, April 1977. (Available from the authors, Office of Research and Evaluation, Austin Independent School District, Austin, Texas).

McCarthy, E.J., Kilpatrick, J.J., and MacNelly, J. A political bestiary: Viable alternatives, impressive mandates, and other fables. New York, NY: Avon, 1978.

McDaniels, G.I. Assessing handicapped students: Beyond identification. New Directions for Testing and Measurement, 1979, 1, 3-8.

McDaniels, G.I. Evaluation problems in Follow Through. In: A.M. Rivlin and P.M. Timpane, (Eds.) Planned variations in education. Washington, D.C.: Brookings, 1975, pp. 47-60.

McKay, H., Sinisterra, L., McKay, A., Gomez, H., and Lloreda, P. Improving cognitive ability in chronically deprived children. Science, 1978, 200, 270-278.

McLaughlin, D.H., Gilmartin, K.J., and Rossi, R.J. Controversies in the evaluation of compensatory education. Palo Alto, CA: American Institutes for Research, July 1977.

McLaughlin, M.W. Evaluation and reform: The elementary and secondary education act of 1965/Title I. Cambridge, MA: Ballinger, 1975.

McLaughlin, M.W. Pygmalion in the school district: Issues for staff development programs. Unpublished paper, Rand Corporation.

Medinal-Swanson, M. Tests and bilingual education. New Directions for Testing and Measurement, 1979, 1, 23-38.

Mielke, K.W., and Swinehart, J.W. Summary: Evaluation of the Feeling Good Television Series. New York, NY: Children's Television Workshop, 1976.

Millman, J., Paisley, W., Rogers, W.T., Sanders, J.R., and Womer, F.B. Performance review of USOE's ESEA Title I Evaluational Technical Assistance Program. Washington, D.C.: Hope Associates, 1979.

Millsap, M.A., National Institute of Education. The uses of evaluation in federal regulation writing. Paper presented at the 1978 meeting of the Evaluation Research Society.

Minor, M.J., & Bradburn, N.M. The Effects of Viewing Feeling Good. Chicago, IL: National Opinion Research Center, University of Chicago, 1976.

Mitchell, D.E. Social science impact on legislative decision making: Process and substance. Paper presented at AERA Annual Convention, Boston, MA, April 1980.

Moles, O.C. Field experiments in preventing disruptive school behavior. Research Memo, National Institute of Education, Washington, D.C., 1977 (Presented at the annual meetings of the American Educational Research Association Meeting, New York, April 1977).

Mosher, F.C. The GAO: The quest for accountability in American government. Boulder, CO: Westview Press, 1979.

Mosteller, F.M. Testimony to the Subcommittee on Science, Research, and Technology, U.S. House of Representatives, on National Science Foundation, February 20, 1980. (Available from the author, Harvard University, Cambridge, MA).

Mosteller, F. and Moynihan, D.P. (Eds.) On equality of educational opportunity. New York, NY: Vintage, 1972.

Murray, C.A. and others. The national evaluation of the Cities in Schools Program. Washington, D.C.: American Institutes for Research, 1979 (NIE Contract No. 400-77-0107).

National Academy of Sciences. Proposal for a study on program evaluation in education to the U.S. Office of Education. Washington, D.C. NAS, July 1979.

National Institute of Education. Administration of compensatory education, Washington, D.C.: NIE, 1977.

National Institute of Education. Compensatory education services, Washington, D.C.: NIE, 1977.

National Institute of Education. Demonstration studies of funds allocation within school districts, Washington, D.C.: NIE, 1977.

National Institute of Education. The effects of services on student development. Washington, D.C.: NIE, September 30, 1977.

National Institute of Education. Evaluating compensatory education: An interim report on the NIE Compensatory Education Study. Washington, D.C.: NIE, 1976.

National Institute of Education, Task Force on Resources Planning. NIE Plan to build educational R&D capacity. Educational Researcher, 1974, 3(2), 14-19.

National Institute of Education. Title I funds allocation: The current formula. Washington, D.C.: NIE, 1977.

National Institute of Education. Using achievement test scores to allocate Title I funds. Washington, D.C.: NIE, 1977.

National Institute of Education. Violent schools--Safe schools: The Safe School Study Report to the Congress. Washington, D.C.: NIE, January 1978.

Newsweek. Basic is better. July 4, 1977, p. 76.

Office of the Assistant Secretary for Planning and Evaluation, Department of Health, Education and Welfare. Report on evaluation utilization in the DHEW, April 1980.

Office of Evaluation and Dissemination, U.S. Office of Education. Briefing book on OED. Washington, D.C.: OED/USOE, 1980.

O'Keefe, M. Letter to Robert Boruch from the Deputy Assistant Secretary, Planning and Evaluation/Education, DHEW, March 20, 1980.

Osterlind, S.J. What makes a school evaluation work? Educational Evaluation and Policy Analysis, 1979, 1(1), 48-51.

Patton, M.Q., et al. In search of impact: An analysis of the utilization of federal health evaluation research. In C.H. Weiss, (Ed.) Using social research in public policy making. Lexington, MA: Lexington Books, 1977.

Pelavin, S.H. and David, J.L. Evaluating long-term achievement: An analysis of longitudinal data from compensatory education programs. SRI Project Report, EPRC 4537-15, Stanford Research Institute, Palo Alto, CA, March 1977.

Pelavin, S.H. Statement on the Fund for the Improvement of the Postsecondary Education, before the Subcommittee on Postsecondary Education, U.S. House of Representatives, July 13, 1979 (based on research supported by Contract HEW-100-78-0128, ASPE).

Pelavin, S.H. Statement on Title I of the Elementary and Secondary School Act. Hearings on H.R. 15, The Subcommittee on Elementary, Secondary, and Vocational Education of the Committee on Education and Labor, U.S. House of Representatives. (Part 19) 95th Congress, First Session, Washington, D.C.: U.S. Government Printing Office, pp. 492-502.

Pancrazio, S.B. State education agencies as research arenas. Educational Researcher, 1978, 7(1), 5-10.

Perez, E. Bilingual education program evaluation in Texas.

Pettigrew, T.F. and Green, R.L. School desegregation in large cities: A critique of the Coleman "white flight" thesis. Harvard Educational Review, 1976, 46, 1-53 (a).

Pickman, J. Memorandum to Rick Cotton, on OS Clearance of OE evaluation summaries and reports, DHEW, Office of Education, Washington, D.C. June 8, 1979.

Pois, J. Watchdog on the Potomac: A study of the Comptroller General of the United States. Washington, D.C.: University Press of America, 1979.

Powell, M. Secondary analysis: In the Beginning Teacher Evaluation Study. New Directions, 1978, 4, 63-74.

Privacy Protection Study Commission. Personal privacy in an information society. Washington, D.C.: U.S. Government Printing Office, 1977.

Program Description, Education Department, Lincoln Hills School, June 1980.

The Providence School District, Evaluation Department. Providence Title I evaluation final report/1977-1978.

The Providence School District, Evaluation Department. Providence Title I evaluation final report/1978-1979.

Raizen, S.A. Dissemination programs at the National Institute of Education. Knowledge Creation, Diffusion, and Utilization, 1979, 1(2), 259-292.

Rand Corporation. Evaluation of the Emergency School Aid Act Non-Profit Organization Program. Santa Monica, CA: Rand Corporation, 1978.

Research Triangle Institute. Evaluation of the Upward Bound Program: A second follow-up. Durham, NC: Research Triangle Institute, 1978.

Rich, R.F. Uses of social science information by federal bureaucrats: Knowledge for action vs. knowledge for understanding. In C.H. Weiss (Ed.) Using social research in public policy making. Lexington, MA: Lexington Books, 1977.

Ridings, J. Hearings report. Unpublished manuscript. Department of Education. Western Michigan University, Kalamazoo, MI, June 14, 1979.

Ridings, J.M. Standard setting in accounting and auditing: Considerations for educational evaluation. Ph.D. Dissertation. Kalamazoo, Western Michigan University, 1980.

Rickel, A.U., Smith, R.L., and Sharp, K.C. Description and evaluation of a preventive mental health program for preschoolers. Journal of Abnormal Child Psychology, 1979, 7(1), 101-112.

Riecken, H.W., Boruch, R.F., et al. Social experimentation. New York, NY: Academic, 1974.

Riecken, H.W. Practice and problems of evaluation: A conference synthesis. In: R.E. Klein, M.S. Read, and H.W. Riecken, (Eds.) The practice of impact evaluation. New York, NY: Plenum, 1979.

Rivlin, A.M., & Timpane, P.M. (Eds.) Planned variation in education. Washington, D.C.: Brookings, 1975.

Rossi, P.H. Booby traps and pitfalls in the evaluation of social action programs. In: F.G. Caro (Ed.) Readings in evaluation research. New York, NY: Russell Sage Foundation, 1977.

Rossi, P.H. Freeman, H.E. and Wright, S.R. Evaluation: A systematic approach. Beverly Hills, CA: Sage, 1979.

Rossi, P.H. Past, present, and future prospects for evaluation research. In L. Datta and R. Perloff, (Eds.) *Improving evaluations*. Beverly Hills, CA: Sage, 1979.

Rost, P. Using evaluation data for program planning in Title I. *Prepost Press*, 1980, *V*(2), 5.

Roupp, R., Travers, J., Glants, F., Coelen, C., and others. *Children at the center: Summary findings and their implications*. Cambridge, MA: Abt, 1979.

Schmidt, R.E., Scanlon, J.W., and Bell, J.B. Evaluability assessment: Making public programs work better. *Human Services Monograph Series*, 1979, 14 (Washington, D.C.: U.S. Government Printing Office, Publication No. 1979-0-305-117; DHEW Publication No. OS-76-130).

Schmidt, R.E., Horst, P., Scanlon, J.W., and Wholey, J.S. *Serving the federal evaluation market*. Washington, D.C.: Urban Institute, 1977.

Scriven, M., and Roth, J. Needs assessment: Concepts and Practice. *New Directions for Program Evaluation*, 1978, Spring (1), 1-12.

Seal, J. Memorandum to Under Secretary and others, regarding departmental policy on the distribution of evaluation summaries and study reports. Office of the Assistant Secretary for Management, Office of the Secretary of Education, Washington, D.C. April 25, 1980.

Searle, B., Friend, J., & Suppes, P. *The Radio Mathematics Project: Nicaragua 1974-75.* Stanford, CA: Institute for Mathematical Studies in the Social Sciences. Stanford University, 1976.

Shoemaker, D.M. Concerning national competency examinations in education. *Educational Evaluation and Policy Analysis*, 1979, *1*(3), 61-70.

Silverman, W.A. The lesson of retrolental fibroplasia. *Scientific American*, 1977, *236*(6), 100-107.

Smith, M.L. Design strategies for experimental studies. In A.M. Rivlin & P.M. Timpane (Eds.) *Planned variations in education*. Washington, D.C.: Brookings, 1975.

Social Science Research Council, Committee on Evaluation. *Report of the Committee*. New York, NY: SSRC, 1978.

Spady, W.G. The impact of school resources on students. In W.H. Sewell, R.M. Hauser, & D.L. Featherman (Eds.) *Schooling and achievement in American society.* New York, NY: Academic Press, 1976.

Sproull, L.S. and Zubrow, D. Standardized testing in the educational organization administrative performance system. Presented at AERA Annual Meeting, Boston, MA, April 1980.

St. Pierre, R.G. The Follow Through debate. *Curriculum Inquiry*, 1979, *9*, 149-161.

Staats, E.B. Letter from the Comptroller General to Senator Russell B. Long and Senator Robert Packwood. (FGMSD-79-48/910360) U.S. General Accounting Office, Washington, D.C., September 25, 1979.

Staats, E.B. Memoranda from the Comptroller General to heads of divisions and offices. Washington, D.C.: U.S. General Accounting Office, April 18, 1980.

Stalford, C. Educational evaluation 1979: A view from NIE Paper presented at the annual meeting of the Evaluation Network. Cincinnati, Ohio, September 24-25, 1979. (Available from the author, National Institute of Education, Washington, D.C.).

Stalford, C. Social structure of third party evaluations. *New Directions for Program Evaluation,* 1980, *5*, 69-81.

State of California Assembly Bill No. 65, 1977.

State of California Assembly Bill No. 551, 1977.

State Program Implementation Studies Branch, Bureau of Education for the Handicapped. *Progress toward a free appropriate education: A report to the Congress on the Implementation of Public Law 94-142.* The Education for all Handicapped Children Act. Washington, D.C.: BEH, DHEW, U.S. Office of Education, January 1979.

State of Illinois, State Board of Education, *Accountability report for Vocational Education in Illinois-FY 1978*. Springfield: SBE, 1978.

State of Kentucky, Bureau of Vocational Education. *Kentucky annual program plan for 1981 and accountability report for 1979*. Lexington, BVE, 1978.

State of Massachusetts, Department of Education, Division of Curriculum and Instruction, Title I, ESEA. Memorandum on interim evaluation reports, from Dr. Richard Zusman to Superintendents of schools and others. Boston, MA: Title I Program Division, DOE, January 4, 1980.

Stebbins, L.B. (Ed.) *Education as experimentation: A planned variation model*. (4 Volumes). Report by the U.S. Office of Education pursuant to Contract No. 300-75-0134. Cambridge, MA: Abt Associates, Inc., 1977.

Stufflebeam, D.L. Handout for a report on a forthcoming publication: Standards for evaluations of educational programs, projects, and materials. Presented at the meeting of the Task Force on Meta evaluation, Annual Convention of the American Educational Research Association, Boston, Massachusetts, April 8, 1980. (Available from the author, Western Michigan University, Kalamazoo, Michigan).

Stufflebeam, D.L. Meta evaluation. Paper presented at the Johns Hopkins University Confrerence on Evaluation, Baltimore, Maryland, November 2, 1979. (Available from the author, Western Michigan University, Kalamazoo, Michigan).

Systems Development Corporation, *National evaluation of ESEA Title I programs in state institutions for the neglected or delinquent, Phase I.* Santa Monica, CA: Systems Development Corporation, 1978/

Tallmadge, G.K. Comments on Dr. Lois-ellin Datta's paper. In *The Education Division's JDRP: Three papers*. New York, NY: EPIE Institute, 1977a. (ERIC #ED156217).

Tallmadge, G.K. *Ideabook: JDRP*. Washington, D.C.: U.S. Government Printing Office, 1977b. (#017-080-01824-1). (ERIC #ED148329).

Tharp, R.G. and Gallimore, R. The ecology of program research and evaluation: A model of evaluation succession. In L. Sechrest, S.G. West, M.A. Phillips, R. Redner, and W. Yeaton (Eds.) *Evaluation Studies Review Annual*, 1979, *4*, 39-60.

Thomas, T.C., Pelavin, S.H. and others. *Patterns in ESEA Title I reading achievement*. (Final Report, SRI Project 4537, OEC Contract 30-76-0025, Assistant Secretary for Education). Menlo Park, CA: SRI International, March 1976.

Tucker, E. Evaluations of the Teacher Corps Program: An overview and critique of evaluation methodologies. Presented at the 1980 Annual Meeting of the American Educational Research Association, Boston, MA, April 9, 1980. (Available from the author, Office of Evaluation, Department of Education, Washington, D.C.).

U.S. Department of Health, Education and Welfare, Office of Education, Office of Evaluation and Dissemination. *Annual evaluation report on programs administered by the U.S. Office of Education: Fiscal Year 1978*. Washington, D.C.: DHEW, USOE, OED, 1979.

U.S. Department of Health, Education and Welfare, Office of Education, Office of Evaluation and Dissemination. *Annual evaluation report on programs administered by the U.S. Office of Education: Fiscal Year 1979*. Washington, D.C.: DHEW, USOE, OED, 1979.

U.S. Department of Health, Education and Welfare. Compendium of HEW evaluation studies. Washington, D.C.: HEW Documentation Center, Office of the Assistant Secretary for Planning and Evaluation, DHEW, Fall, 1977.

U.S. General Accounting Office. An assessment of the Department of Housing and Urban Development's Experimental Housing Allowance Program. Washington, D.C.: GAO, 1978 (CED-78-29).

U.S. General Accounting Office. Assessing social program impact evaluations: A checklist approach. Washington, D.C.: U.S. General Accounting Office, October 1978. (PAD-79-2).

U.S. General Accounting Office. Evaluation and analysis to support decision-making. Washington, D.C.: U.S. GAO, 1975 (PAD-76-9).

U.S. General Accounting Office. Exposure Draft. Evaluating programs with multiple and conflicting objectives: A framework and some methods. Washington, D.C.: U.S. GAO, February 1980 (PAD-80);

U.S. General Accounting Office. Federal program evaluations: A directory for the Congress (Congressional Sourcebook Series). Washington, D.C.: U.S. Government Printing Office, 1976.

U.S. General Accounting Office. Follow Through: Lessons learned from its evaluation and the need to improve its administration. Washington, D.C.: U.S. GAO, 1975 (MWD-75-34).

U.S. General Accounting Office. Problems and needed improvements in evaluating Office of Education programs. Washington, D.C.: U.S. GAO, September 8, 1977 (HRD-76-165).

U.S. General Accounting Office. Finding out how programs are working: Suggestions for Congressional oversight. Washington, D.C.: U.S. GAO, November 3, 1977 (PAD-78-3).

U.S. General Accounting Office, Program Analysis Division. Background paper for use by the SSRC Committee on Evaluation of the need for access by GAO auditors in the audit of social research and experiments. (Mimeo) Washington, D.C.: U.S. GAO, PAD, April 8, 1978.

U.S. General Accounting Office. Progress in improving program and budget information for Congressional use. Washington, D.C.: U.S. GAO (PAD-79-79) August 31, 1979.

U.S. General Accounting Office. Status and Issues: Federal program evaluation. Washington, D.C.: U.S. GAO, October 1978.

U.S. General Accounting Office. What is the role of federal assistance for vocational education. Washington, D.C.: U.S. GAO, 1974.

U.S. House of Representatives, Committee on Appropriations. *Hearings on DHEW Appropriations*. Washington, D.C.: U.S. Government Printing Office, 1979.

U.S. House of Representatives, Committee on Appropriations. *Report on Departments of Labor and Health, Education and Welfare, and related agencies appropriations bill, 1978*. Washington, D.C.: U.S. Government Printing Office, 1977.

U.S. House of Representatives, Committee on Appropriations. *Report on Departments of Labor and Health, Education and Welfare and related agencies appropriations bill, 1979*. Washington, D.C.: U.S. Government Printing Office, 1978.

U.S. House of Representatives, Committee on Appropriations. *Report on Departments of Labor and Health, Education and Welfare, and related agencies appropriations bill, 1980*. Washington, D.C.: U.S. Government Printing Office, 1979.

U.S. House of Representatives, Committee on Education and Labor, Subcommittee on Elementary, Secondary and Vocational Education. *Hearing on H.R. 15 Part 18: Administration of Title I of ESEA.* (95th Congress, First Session). Washington, D.C.: U.S. Government Printing Office, October 25, 26, 28, 1977.

U.S. House of Representatives Committee on Education and Labor. *Hearings on the Youth Employment Act*. Washington, D.C.: U.S. Government Printing Office, 1979.

U.S. House of Representatives, Committee on Education and Labor, Subcommittee on Elementary, Secondary, and Vocational Education. *Oversight hearing on Title I of the Elementary and Secondary Educational Act*. (95th Congress, First Session). Washington, D.C.: U.S. Government Printing Office, March 6, 1979.

U.S. House of Representatives, Committee on Education and Labor. *Oversight Hearings of Vocational Education*. Washington, D.C.: U.S. Government Printing Office, 1979.

U.S. House of Representatives, Committee on Education and Labor. *Report: The Education Amendments of 1978, H.R. 15.* (95th Congress, Second Session Report No. 95-1137). Washington, D.C.: U.S. Government Printing Office, May 11, 1978.

U.S. House of Representatives, Committee on Education and Labor, *Report: The Education Amendments of 1980*. (96th Congress, First Session, Report 95-520) Washington, D.C.: U.S. Government Printing Office, October 17, 1979.

U.S. Office of Education, Education Programs that work. San Francisco, CA: Far West Laboratory, 1979.

U.S. Office of Education, Office of Career Education. How to get JDRP approval of career education projects. Mimeo, 1980.

U.S. Office of Education, Office of Evaluation and Dissemination. Executive summary: Assessment of programs and projects funded under Title IV of PL 92-318, Indian Education Act, Part A. Washington, D.C.: USOE, 1978.

U.S. Office of Education, Office of Evaluation and Dissemination. Executive summary: Providing compensatory services. Washington, D.C.: USOE, 1978.

U.S. Office of Education, Office of Evaluation and Dissemination. Impact of evaluation studies on OE programs. Mimeo.

U.S. Office of Education, Office of Evaluation and Dissemination. Impact of evaluation studies on OE Programs: National Diffusion Program.

U.S. Office of Education, Office of Evaluation and Dissemination. National Diffusion Network Policies: Adoption Definition. January 31, 1979, Mimeo.

U.S. Office of Education, Office of Planning, Budgeting, and Evaluation. Executive Summary. Conditions and process of effective school desegregation. Washington, D.C.: USOE/OPBE, September 1976.

U.S. Office of Education, Office of Planning, Budgeting and Evaluation. Executive Summary. Elementary schools and the receipt of compensatory funds. Washington, D.C.: USOE/OPBE, August 1977.

U.S. Office of Education, Office of Program Planning, Budgeting, and Evaluation. Executive Summary. Evaluation of Federal programs supporting educational change. Washington, D.C.: USOE/OPBE, October 1977.

U.S. Office of Education, Office of Program Planning and Evaluation. Executive Summary. Evaluation of the field test of Project Information Packages. Washington, D.C.: USOE/OPBE, October 1977.

U.S. Office of Education, Office of Planning, Budgeting, and Evaluation. Executive Summary. Final report on the evaluation of the impact of ESEA Title VII Spanish/English Bilingual education program. Washington, D.C.: USOE, March 1978.

U.S. Office of Education, Office of Program Planning, Budgeting and Evaluation. Executive Summary. The identification and description of exemplary bilingual programs. Washington, D.C.: USOE/OPBE, January 1976.

U.S. Office of Education, Office of Planning, Budgeting, and Evaluation. *Executive Summary. Secondary analysis of the Emergency School Assistance Act Program.*

U.S. Office of Education, Office of Planning, Budgeting, and Evaluation. *Executive Summary. Student economic background achievement status and selection for compensatory services*. Washington, D.C.: USOE/OPBE, April 1978.

U.S. Office of Education, Office of Program, Planning, and Evaluation. *Executive Summary. Title I of the Elementary and Secondary Education Act of 1965 in state institutions for the neglected or delinquent/a national evaluation, Phase I.* Washington, D.C.: USOE/OPBE, October 1977.

U.S. Office of Education, Office of Planning, Budgeting and Evaluation. *Executive Summary. Title IV of the Civil Rights Act of 1964: A review of program operations*. Washington, D.C.: UOSE/OPBE, November 1976.

U.S. Office of Management and Budget. *Management improvement and the use of evaluation in the executive branch.* Circular No. A-117. Washington, D.C.: U.S. OMB, March 23, 1979.

U.S. Office of Management and Budget. *Resources for program evaluation*. Washington, D.C.: U.S. OMB, 1977.

U.S. Senate, Committee on Appropriations. *Report: The Departments of Labor and Health, Education, and Welfare, and Related Agencies Appropriation Bill, 1980, HR 1489*. (96th Congress, First Session, Report No. 96-247) Washington, D.C.: U.S. Government Printing Office, July 13, 1979.

U.S. Senate, Committee on Human Resources. *Report: Education Amendments of 1978, S. 1753* (95th Congress, Second Session, Report No. 95-856). Washington, D.C.: U.S. Government Printing Office, May 15, 1978.

U.S. Senate Committee on Labor and Human Resources. *Hearings on cost, management, and utilization of human resources program evaluations*. 95th Congress. Washington, D.C.: U.S. Government Printing Office, 1978.

U.S. Senate, Subcommittee on Oversight Procedures, Committee on Government Operations. *Legislative oversight and program evaluation: Seminar sponsored by the Congressional Research Service*, May, 1976. Washington, D.C.: U.S. Government Printing Office, 1976.

Wang, Ming-mei. Evaluating the effectiveness of compensatory education. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, April 7-11, 1980. (Available from the author, System Development Corporation, SES Group, Santa Monica, CA).

Wargo, M.J., Campeau, P.L., and Tallmadge, G.K. Further examination of exemplary programs for educating disadvantaged children. American Institute for Research, Palo Alto, CA, July 1971. Contract number OEC-0-70-5016 from USOE, DHEW.

Webster, W.J., & Holley, F. A position paper and recommendations proposed by the Texas Joint Urban Evaluation Council for a state-supported research and evaluation system for Texas school districts, 1974.

Webster, W.J. and Stufflebeam, D.L. The state of theory and practice in educational evaluation in large urban school districts. Unpublished manuscript. Western Michigan University, Kalamazoo, Michigan, 1980.

Weikart, D.P. & Banet, B.A. Model design problems in Follow Through. In: A.M. Rivlin and P.M. Timpane (Eds.) Planned variations in education. Washington, D.C.: Brookings, 1975, pp. 61-77.

Weinberg, M. ESAA: Digging deeper into finances. Journal of Educational Statistics, 1978, 3(1), 95-97.

Weinberg, H. The role of cognitive biases in the interpretation of program evaluations. Unpublished manuscript. Arlington, Virginia, 1979.

Weiberg, H. Using policy analysis in Congressional Budgeting. In F.M. Zweig, (Ed.) Evaluation in legislation. Beverly Hills, CA: Sage, 1979, pp. 28-44.

Weiss, C. Where politics and evaluation meet. Evaluation, 1973, 1(3), 37-45.

Weiss, C.H., and Bucuvalas, M.J. Truth tests and utility tests. American Sociological Review, 1980, 45, 302-312.

Weissart, W.G., Wan, T.T.H., and Livieratos, B.B. Effects and costs of day care and homemaker services for the chronically ill: A randomized experiment. NCHSR Research Report Series. Washington, D.C.: Department of Health, Education and Welfare, National Center for Health Services Research, February 1980. (DHEW Publication No. (PHS) 79-3258).

Weissman, C. Overview of the Abt Associates, Inc. National Evaluation of Follow Through. Hearings before the Senate Subcommittee on Employment, Poverty, and Migratory Labor of the Senate Human Resources Committee, for the Economic Opportunity Amendments of 1978. May 14, 1978, pp. 860-871.

Welch, W.W., & Walberg, H.J. A national experiment in curriculum evaluation. American Educational Research Journal, 1972, 9, 373-383.

APPENDICES 1-4

PROJECT ON EVALUATION OF EVALUATIONS OF FEDERAL EDUCATION PROGRAMS

Northwestern University
Evanston, Illinois

Appendix 1.	Legislative and Management Background of the Project: Excerpts from Public Law 95-561, Conference Report, Congressional Record, and USOE Work Statement
Appendix 2.	Principal Staff of the Project on Educational Program Evaluations and their Responsibilities
Appendix 3.	Research Strategy and Methods
Appendix 4.	Tabular Comparison of Standards for Evaluation

APPENDIX 1

Legislative and Management Background of the Project:
Excerpts From Public Law
95-561, Conference Report,
Congressional Record, and USOE Work Statement.

Section 1526. The Commissioner of Education shall conduct a study of evaluation practices and procedures at the national, state, and local levels with respect to federally funded elementary and secondary educational programs and shall include in the first annual report to Congress submitted more than one year after the data of enactment of this Act proposals and recommendations for the revision or modification of any part or all of such practices and procedures. Such proposals and recommendations shall include provisions--

(1) to ensure that evaluations are based on uniform methods and measurements;

(2) to ensure the integrity and independence of the evaluation process; and

(3) to ensure appropriate follow-up on the evaluations that are conducted.

EXCERPT: CONFERENCE REPORT

32. Study of Evaluation

The House bill, but not the Senate amendment, requires the Commissioner to conduct a comprehensive study of evaluation practices and procedures at the national, state and local levels with respect to federally funded elementary and secondary programs, and to submit a report within 1 year.

The Senate recedes with an amendment encouraging this study to be intensive on certain problems instead of comprehensive.

EXCERPT: CONGRESSIONAL RECORD
HOUSE, H6671, 1978

Amendment offered by Ms. Holtzman

Ms. HOLTZMAN, Mr. Chairman, I offer an amendment.

The Clerk read as follows:

Amendment offered by Ms. HOLTZMAN: Page 375. Insert the following new section after line 25:

COMPREHENSIVE STUDY OF EVALUATION
Practices and Procedures

"Sec. 1331. The Commissioner of Education shall conduct a comprehensive study of evaluation practices and procedures at the national, state, and local levels with respect to federally-funded elementary and secondary educational programs and shall report to Congress within one year after the date of enactment of this Act with proposals and recommendations for the revisions and modification of any part or all of such practices and procedures. Such proposals and recommendations shall include provisions-

(1) to ensure that evaluations are based on uniform methods and measurements;

(2) to ensure that integrity and independence of the evaluation process; and

(3) to ensure appropriate follow-up on the evaluations that are conducted."

Redesignate the following section and conform the table of contents accordingly.

Ms. HOLTZMAN. (during the reading). Mr. Chairman, I ask unanimous consent that the amendment be considered as read and printed in the RECORD.

The CHAIRMAN. Is there objection to the request of the gentlewoman from New York?

There was no objection.

(Ms. HOLTZMAN asked and was given permission to revise and extend her remarks.)

Ms. HOLTZMAN. Mr. Chairman, I want to compliment the Committee on Education and Labor and especially the distinguished and able Chairman of the committee (MR. PERKINS) for this very important bill on elementary and secondary education.

I am offering a modest but, I think, necessary and useful amendment. It simply would require the Commissioner of Education to review the present procedures for evaluation to see how they could be improved and to report back to Congress with his suggestions and recommendations within 1 year.

I am concerned that we try to get the most out of our education dollars. This is impossible without an effective evaluation process which allows us to learn from our mistakes and build on our successes.

At present, most federally funded programs in elementary and secondary schools must be evaluated, but the evaluation process is inadequate. There are no requirements for follow-up on evaluations-and many, including highly critical ones, are filed away or forgotten. The absence of any uniform procedures or standards makes it difficult to compare different education programs and permits the use of inadequate evaluations at the local and state level. Another problem is that many school districts select their own evaluators, who in order to obtain future contracts may be less than candid in their appraisals.

To improve education programs, we need honest, thorough, effective evaluations which are followed up.

My amendment calls on the Committee of Education to recommend ways of improving the integrity and effectiveness of evaluations. Without such evaluations we cannot be sure that our education dollars are as well spent as they should be.

Mr. PERKINS. Mr. Chairman, will the gentlewoman, Ms. HOLTZMAN, yield to me?

Ms. HOLTZMAN. I would be delighted to yield to the distinguished gentleman from Kentucky.

Mr. PERKINS. I thank the gentlewoman for yielding.

Let me say that in my judgment the gentlewoman's amendment is entirely appropriate and fitting at this palce in the RECORD. It is something that should be required, appropriate evaluations, and we accept the amendment.

Mr. QUIE. Mr. Chairman, will the gentlewoman yield?

Ms. HOLTZMAN. I yield to the gentleman from Minnesota.

Mr. QUIE. I thank the gentlewoman for yielding.

I believe that the 1-year comprehensive study of the evaluation practices would be very beneficial to this committee, and I would be happy to accept the amendment.

Ms. HOLTZMAN. I thank the gentlemen for their comments.

Mr. Chairman, I yield back the remainder of my time.

The CHAIRMAN. The question is on the amendment offered by the gentlewoman from New York. (Ms. HOLTZMAN)

The amendment was agreed to.

APPENDIX 1

EXCERPT: WORK STATEMENT
ISSUED BY THE U.S. OFFICE OF EDUCATION
OFFICE OF EVALUATION & DISSEMINATION
IN RESPONSE TO SECTION 1526, PUBLIC LAW 95-561

1. The Investigator. In view of the self-evaluation aspect of the study, it should not be carried out by those persons or groups ordinarily involved in the evaluation of Federally funded education programs. The services of a prestigious and experienced person with a national reputation in educational evaluation who is known for independence and impartiality should be obtained to oversee the study and to report directly to the Commissioner.

2. The Questions to Be Asked. The questions posed by Rep. Holtzman's request should be expanded upon in a more detailed fashion and priority should be assigned to them, especially with regard to the specificity with which they will be addressed at each of the three levels of the study--local, state and federal. This should be done in part by conferring directly with Rep. Holtzman. The three main questions encompass a number of allied questions that can be addressed. Examples of some of these are given below:

(a) Why and how are evaluations carried out (methods and measurements)?

- (i) for whose information needs are evaluations carried out?
- (ii) what procedures are used in carrying them out?
  - (iia) how "sound" are they (accurate, valid, reliable, etc.)?
- (iii) what measurement techniques and devices are used and, how appropriate are they to the program's objectives?
- (iv) are there conflicting requirements for different programs that lead to duplicative or unusually burdensome efforts?

(b) What are the capabilities (including integrity and independence) of those who carry out evaluations?

- (i) where are evaluation activities located organizationally?
- (ii) what is the background, training and experience of the staff and of those who carry out evaluations? If such services are provided by outside agencies, how are their services obtained and what is the nature of their relationship to in-house staff?

(c) how are the results of evaluations used?

(i) what are the conditions that facilitate or detract from their use?

(ii) when results are used, what is the nature of the changes they lead to?; can examplary uses be identified?; is their nature such that they can be adopted in other settings?

(d) what legislative regulatory, funding or other changes might be proposed as means to improving the nature, conduct and utility of evaluations?

3. The Programs to Be Examined. Clearly, not all federally funded elementary and secondary programs can or should be examined. Rather a set of programs that are numerous enough and sufficiently diverse enough (e.g., State administered versus direct grant) to allow one to tease out all the issues involved yet are few enough in number to allow an intensive examination of certain problems (as suggested by the Senate in the conference report) should be included. Examples of these are: Title I and VII, ESEA; ESAA; Handicapped; and Vocational Education (these are programs that have a local evaluation requirement; two are direct grant programs (ESAA and Title VII) while the remainder are State administered).

4. The Conduct of the Study. The investigator would prepare a detailed Work Statement and would acquire the services of an 8a contractor to provide him with the administrative, logistical and personnel support needed to:

(a) Review reports of those agencies involved in evaluation.

(b) Interview persons involved in the initiation, conduct and utilization of evaluations,

(c) Convene a panel of prestigious persons to review the findings from (a) and (b) and to make recommendations for the improvement of evaluation practices.

APPENDIX 2.

PRINCIPAL STAFF OF THE PROJECT ON EDUCATIONAL PROGRAM EVALUATIONS AND THEIR RESPONSIBILITIES

(Principal staff are indicated by an ASTERISK)

*Robert F. Boruch - Principal Investigator. Boruch is Director of the Division of Methodology and Evaluation Research (DMER), Professor of Psychology at Northwestern University, and Director of the Project. His research, scholarly papers, and books concern the development of methods to planning and evaluating social programs. He holds a faculty appointment in the Center for Applied Probability and Statistics and is Chairman of the Social Science Research Council's Committee on Program Evaluation.

*David S. Cordray - Co-Principal Investigator. Dr. Cordray is Associate Director of the Division of Methodology and Evaluation Research (DMER) and Assistant Professor of Psychology at Northwestern University. His previous experience in education, juvenile justice, and child health services has included local, state and federal level evaluation efforts.

Joe S. Cecil - Legal-Legislative Issues. Cecil is an attorney with a Ph.D. in methodology and evaluation research. As a post-doctoral Fellow at Northwestern University, he was responsible for legal research at the federal level. Cecil is currently at the Federal Judicial Center.

*Stephany Creamer - Systems Manager. Creamer has a master's degree in Educational Administration and has been an elementary school teacher, a member of the staff of the Office of Research of the Chicago School System and staff of the Board of Education of Chicago. She has been responsible for coordination of site visits, and round-tables, and for conducting site visits.

Leslie Goldgehn - Research Assistant. Goldgehn has an MEd from Loyola University in Higher Education Administration and Policy.

*Laura Leviton - Utilization/Dissemination Team Leader. Dr. Leviton is a post-doctoral Fellow in the Division of Methodology and Evaluation Research at Northwestern. She has done research on utilization of evaluations and on diffusion of research findings for the Health Care Finance Administration, and was responsible for similar research in this project.

Charles Gene-McDaniel - Editor and Science Writer. McDaniel, an award-winning writer and editor for the Chicago Bureau of Associated Press for 20 years, is currently a member of the faculty at Roosevelt University.

Lynne Miller - Trainer and Consultant. Dr. Miller, a specialist in organization of schools, has been a teacher in urban Philadelphia schools, a member of the Amherst faculty, and a member of the Task Force on Utilization of Research in the Schools.

Robert Orwin - Research Assistant. Orwin is a Ph.D. candidate in Methodology and Evaluation Research, a principal interviewer on LEA site visits, and a technical assistant in development of interview protocols.

*Georgine Pion - Manpower Capabilities Team Leader. Dr. Pion is a social psychologist. As a post-doctoral Fellow at Northwestern University, she has been responsible for assessing issues related to manpower capabilities, supply, and organizational consideration relevant to hiring and maintaining quality evaluation staff.

Jerry Ross - Organizational Characteristics Team Leader. Dr. Ross is an Assistant Professor of Organizational Behavior in the Kellogg Graduate School of Management, Northwestern University. His training is in management and accounting. His primary area of responsibility has been investigating organizational issues that effect evaluation practices.

Barbara Schneider - Research Associate and Consultant. Dr. Schneider is Assistant Dean for Research in the School of Education at Northwestern. She specializes in research on the economics of Education. She took responsibility for providing some general information and for some cost computations.

Roger Straw - Research Assistant. Straw is a Ph.D. candidate in Methodology and Evaluation Research. He has been responsible for review of evaluation standards and guidelines and investigations of utilization research.

Sara L. Thomas - Research Assistant. Thomas is a Ph.D. candidate in Methodology at Northwestern. Her primary duties have been organizational execution of telephone surveys.

William Trochim - Research Associate. Trochim has been responsible primarily for eliciting information from Technical Assistance Centers on evaluation designs used by Center clients. He is currently Assistant Professor at Cornell University.

Janet Weeks - Research Associate. Weeks, a former high school teacher, is a Ph.D. candidate in the School of Education at Northwestern, and has participated in evaluations of CETA Youth Employment Programs. She has been responsible for obtaining information on vocational education.

John Wick - Bilingual Education Specialist. Dr. Wick, formerly Director of Research for the Chicago School District, is currently a professor of Education at Northwestern University. He provided expert advice and assistance on LEA and SEA evaluations and site visits.

Penny Wohlstetter - Research Assistant. Wohlstetter, with a Masters in Education and Public Policy from Harvard, has had responsibility for document acquisition and report synthesis. She has worked as a policy analyst with the U.S. Office of Education.

Carlotta Young - Research Associate. Young has been responsible for field research on utilization of evaluation results. She had been an intern at the U.S. Accounting Office before becoming involved in this Project and received her Ph.D. from Pennsylvania State University. She is currently with the U.S. General Accounting Office.

PRINCIPAL AUTHORS AND EDITORS

Robert Boruch and David Cordray have had primary responsibility for development and editing of the Project's final report. Principal authors of major subsections include Georgine Pion on evaluator capabilities, Laura Leviton on utilization, Ceceil on law, Ross on organizations, and Roger Straw on standards. Case Study material was written by almost all members of the Project staff, especially Wick, Creamer, Orwin, Weeks, Boruch, Pion, Leviton, and Cordray.

CONSULTANTS

Participants in each round-table discussion were treated as consultants in this Project. They are listed in Appendix 3 and include representatives of school boards, and of local and state education agencies, former and current Congressional staff members, university faculty, and senior researchers in contract research groups.

APPENDIX 3. RESEARCH STRATEGY AND METHODS

CONTENTS

Page

3.1 Introduction 1
3.2 Recent Research on Evaluation: Abstracts 3
3.3 Site Visits: Federal Agencies 9
3.4 Site Visits and Telephone Surveys: Local and State Education Agencies 10
3.5 Site Visits and Telephone Interviews: Title I Models . . . 17
3.6 Round-table Discussions 18
3.7 Formal Presentations and Discussions 22
3.8 Machine Based Searches 23
3.9 Evasive Responses and Tactics 25

3.1 Introduction

The main justification for this Project is Congressional interest in the character of federally supported program evaluations. In particular, Section 1526 of Public Law 95-561 requires that "The Commissioner of Education shall conduct a study of evaluation practices and procedures at the national, state, and local levels with respect to federally funded elementary and secondary educational programs..."

The study design was based on discussion among Congressional staff, federal agency personnel, and Project staff. The study was undertaken specifically to furnish appraisal independent of federal agencies and to examine evaluation at federal, state, and local levels of government. The Project is prospective in orientation, designed to provide: recommendations about evaluation policy and practice, the evidence to sustain recommendations, and the identification of issues and options. Specific questions to be addressed by the Project include:

- Why and how are evaluations carried out?
- What are the capabilities of those who carry out evaluations?
- How are the results of evaluation used?
- What recommendations can be made to improve policy or practice?

Answers to each question were obtained for federal, state, and local levels of administration, as requested by the Congress. The main vehicles for providing answers to such questions are a critical examination of contemporary research on each topic, field work by Project staff, and roundtable discussions and formal presentations at professional meetings.

Contemporary research that was reviewed by the Study staff included major evaluation studies conducted by HEW's Division of Education, special studies of the quality and uses of evaluation data at local, state, and federal level, and Congressional testimony and records on initiation conduct and use of evaluations. The special studies have been supported by the National Institute of Education, U.S. Office of Education, and National Science Foundation. Samples of evaluations undertaken at the local school district level were also reviewed. This part of our investigation was supplemented by interviews with the individuals responsible for production of the research.

The second major source of information was interviews by Project staff directly from individuals with an interest in evaluation of educational programs. Three groups were involved. The first was local school district staff who were interviewed both in site visits and in a telephone survey. The second was state officials with responsibility in the area, again interviewed on site and through telephone surveys. The final group included federal agency staff and Congressional staff, interviewed primarily through site visits.

APPENDIX 3

The third major source of information was roundtable discussions conducted at Northwestern University, and discussion of Project activities at professional society meetings. The topics for each roundtable discussion included utilization, local school evaluation, and parental interests in evaluation, and others.

Appendix 3.2

Recent Research on Evaluation:
Abstracts and Literature Review

Project on Evaluations of Federally Funded Education Programs relied heavily on earlier research by Northwestern and by other private and public institutions. The major studies on which we have relied are summarized below. The ones most relevant to evaluation capabilities and organization include:

- UCLA Center for the Study of Evaluation, Survey of Large School District Evaluation Units
- Bureau of Social Science Research, Survey of Performers of Research and Research Related Activities
- Northwestern Project on Evaluation of Evaluations
- Hope Associates Performance Review of Technical Assistance Centers

The major resources for information on use of evaluation results and of research more generally include studies at local, state, and federal levels. For the local level, these include:

- SRI International, Evaluation of the National Diffusion Network
- SRI International, Study of Local Uses of Title I Evaluation
- UCLA Center for the Study of Evaluation, Survey of School District Evaluation Units (Lyons and others) and Case Studies (Alkin and others)
- Rand Change Agent Study and Datta (NIE) Reanalysis
- Huron Institute, Study of Local Uses of Evaluation (unavailable at this writing).

The state level studies include:

- SRI International Evaluations of the National Diffusion Network
- Bissell Case (California) Study

Federal level studies include:

- Florio (AERA) Survey of Congressional Staff
- Milsap (NIE) Case Study
- DHEW/ASPE (Wholey) Survey

Studies of Capabilities and Organization

The following material summarizes each of these studies. References to reports are furnished at the end of this appendix.

UCLA Center for the Study of Evaluation's Survey of Large School District Evaluation Units

During 1977-78, the University of California's Center for the Study of Evaluation examined the organization of local school district offices of evaluation. The Center is a regional laboratory, supported by the National Institute of Education. This study focused on a target population of school districts having an enrollment of 10,000 or more students and an organizational unit for evaluation. Some 350 districts which met both of these criteria were surveyed and 72% responded.

The study obtained data on the number of districts with evaluation units, the qualifications of unit directors, size of staff and staff needs, activities falling under the rubric of evaluation, organizational characteristics of units, use of consultants, and funding. Details are given in a report Evaluation and School Districts by Lyon, Doscher, McGranahan, and Williams. The study was a major resource for Northwestern's examination of definition, capabilities, and organization of evaluation at the local level.

Bureau of Social Science Research's Survey of Performers of Research and Research Related Activities

During 1976-78, the Bureau of Social Science Research (BSSR), a private contracting firm in Washington, D.C., undertook a study to identify and describe nonfederal organizations that conduct research, development, dissemination, and evluation (RDD&E) in education. Screening of over 6,000 organizations resulted in a group of 2,434 organizations within 1,530 institutions which provided information on capabilities. Screening was based on the organization's having been an active performer of educational RDD&E during the preceding year, having a distinct organizational identity, and having appreciable autonomy in carrying out educational RDD&E. The largest subgroup of institutions in the respondent group was the public education sector: 37 state education agencies, 193 intermediate service agencies, 401 local education agencies. The academic sector, colleges and universities, included 423 institutions. There were 476 miscellaneous institutions such as private contractors. BSSR obtained information on funding and expenditures of the organizations, staff size and education level, functional emphases, number, duration, and character of projects. Details are given in a report by Frankel, Sharp, and Biderman.

Northwestern University's Project on Evaluation of Evaluations

The Northwestern field study has been dedicated to confirming some findings of the earlier studies, examining more closely arguable findings, based on survey of school districts, state education agencies, and federal agencies. This includes identifying staff background, training, major responsibilities, use of outside contractors, and difficulties in hiring competent staff and upgrading skills. It also includes reanalysis of earlier studies by the Bureau of Social Science Research and UCLA's Center for the Study of Evaluation, and identifying issues which were not addressed in the earlier work.

Hope Associates, Inc.'s Performance Review of Technical Assistance Centers

Ten Technical Assistance Centers (TAC) have been operating since 1976 to provide aid to state and local education agencies involved in Title I program evaluations. They were reviewed during 1978-79 by a panel of educational experts under contract to Hope Associates, with DHEW's Office of Assistant Secretary for Planning and Evaluation. The review involved site visits to each TAC, to a dozen federal agency executives, and to 25 state education agencies, telephone surveys of the remaining 25, and examination of pertinent printed material. Details are given in a report issued by Hope Associates.

Studies of Utilization

Stanford Reserach Institute's Evaluation of the National Diffusion Network

The National Diffusion Network (NDN) was established by USOE in 1974 to foster diffusion and adoption of exemplary education programs. It was evaluated during 1975-76 to "understand the relative effectiveness" of the system. The evaluation was done by SRI under contract with the USOE's Office of Planning Budgeting and Evaluation. Evaluation involved analysis of pertinent source documents, mail survey of over 900 local education agencies, over 40 program developers, and over 60 facilitators, 16 developers, and 36 adopters, site visit interviews with 149 teachers, and 30 principals. The main goals of the evaluation were to provide comprehensive description of the NDN process, to evaluate the organizational effectiveness of NDN, and to make recommendations based on the inquiry. We are aware of no critique or secondary analysis. Details are given in a report by Emrick and others.

SRI International's Study of Local Uses of Title I Evaluations

The study, supported by the Office of Assistant Secretary for Planning and Evaluation, was designed to determine whether local school district staff used Title I evaluation results to "identify strengths and weaknesses of their programs in order to improve them" and "whether the recent and proposed changes in Title I evaluation was likely to alter local use of evaluation"

(p. iii). The study focused on 15 Title I districts in three states, selected through referrals of recommendations by federal staff and Technical Assistance Directors on the basis of their beliefs that the districts were "especially concerned with evaluation." Site visits were undertaken to interview administrators, principals, teaching staff, and parents. Details are given in a 1978 report by Jane David.

UCLA Center for the Study of Evaluation's (Alkin, Daillak, and White) Case Studies in Compensatory Education and Innovative Projects

Five in-depth case studies examined the uses of evaluations in projects for Title I compensatory education and innovations at the local level (Title IV-C of ESEA). The studies described the community, school district and setting of the project, the decision makers and evaluators, the nature of the project, the nature of the evaluation, and uses made of the evaluation. In all cases, at least some use was made of evaluation information. Different audiences used evaluations differently. The evaluation frequently affected thinking and decisions jointly with other information. Details are given in a 1979 report by Alkin, Daillak, and White.

UCLA Center for the Study of Evaluation's (Resnick, O'Reilley, and Majchrzak) Survey of School Districts

Using the survey data collected in the Center for the Study of Evaluation's survey of evaluation units within school districts, these researchers concluded that evaluations geared toward resource allocation decisions were used more by the superintendent and the school board, while evaluations directed at developing or modifying curriculum were used by the program administrator, with a trend toward use by teachers and the superintendent. Details are presented in a report by Resnick and others.

UCLA's Center for the Study of Evaluation's (Alkin, et al) Study of Title VII

Evaluation reports and independent audits of the reports were examined for 42 projects in bilingual education funded under Title VII. The authors also obtained federal monitors' ratings of the quality of projects and questionnaire responses from the project directors. Although the authors stress that their data are not stable, with extremely wide confidence limits, a sample of the findings is presented. See the 1974 Alkin et al report for details.

RAND's Study of Federal Programs Supporting Educational Change

Implementation and continuation of programs aimed at inducing educational change in local school districts were examined. These involved four federal programs: Title III of ESEA, the Elementary and Secondary Education Act; Right to Read, Vocational Education Act Part D; and ESEA Title VII (bilingual). Two hundred ninety three projects were studied in

18 states, involving interviews with teachers, principals, program managers, superintendents, and federal program administrators. Twenty-nine projects were studied in depth. One hundred projects were reexamined to determine whether projects were continued at the end of federal funding. Implementation and continuation of projects were affected by: 1) the degree to which the innovation matched the local education agencies objectives; 2) the degree to which the innovation was consonant with the values of the local education agency; amount of change required, and 3) complexity of the innovation. See the report by Berman and McLaughlin for details.

Datta's Reanalysis of the RAND Study

Lois-ellin Datta of the National Institute of Education challenged the conclusion of RAND's report, "Federal Programs Supporting Educational Change" that use of outside technical assistance was not effective in promoting local school district change. First, few characterizations of such experts could be found in the reports. Second, a follow-up questionnaire indicates greater use of experts than appears in RAND's conclusions. All forms of assistance, local and outside, were viewed as being not very useful by teachers. Outsiders were perceived as being more useful in those studies in which program goals were achieved. In most instances, teacher participation and involvement was not more strongly related to project success than use of outside experts. Datta therefore believes that this particular conclusion of the RAND report should no longer be cited. See Datta's 1980 paper for detail.

Bissell's Case Study

Joan Bissell presents four cases in which the California state legislature used evaluations in making decisions about the form and funding for programs. She also discusses use of basic indicators by the legislature, and offers suggestions for improving use of evaluations by legislatures.

Florio's Survey of Congressional Staff

Twenty-six Congressional staff members who dealt with education issues were interviewed. Agencies serving Congress were the most frequent source of research information, and in general, Washington based sources were relied on most. Cost information was most frequently mnetioned as useful, followed by achievement scores. Different types of information were useful at different phases of the congressional cycle. Evaluations compete with other information for Congressional attention and trust. Details are given in a paper by David Florio.

Mitchell's Study of State Legislators

Mitchell interviewed 160 legislators and staff members in the states of Arizona, California and Oregon about issues in education. Respondents described legislation in education in recent years, the situation and issues involved in this legislation, and the resources used to influence decision making. Mitchell's results point to the importance of legislative cycles in getting information used.

Milsap's Case History of Experience Based Career Education

A case history is presented of the manner in which the Experience Based Career Education program became the priority program for funding and adoption through Vocational Education Act, Part D. The program and its evaluation are described, its process through the Joint Dissemination Review Panel is given, and the process by which regulations were revised so that the program would have priority. Milsap raises two issues in this case history of research utilization. The first relates to the lack of visibility of regulations, and the degree to which the public becomes involved. The second relates to the adequacy and scope of evaluations that are intended to be used for such purposes and to potential issues of misutilization. See Milsap's 1979 report for details.

Literature Review

A great deal of the literature bearing on evaluation of federally funded programs was reviewed *before* this Project was undertaken. Earlier work supported by the National Institute of Education was published as:

Boruch, R. F., and Wortman, P. M. Implications of educational evaluation for evaluation policy. In: D. C. Berliner (Ed.) *Review of Research in Education*. Washington, D.C.: American Educational Research Association, 1979, pp. 309-363.

Earlier work supported by the Health Care Finances Administration has been issued as:

Leviton, L. C., and Hughes, E. F. X. Utilization of evaluations: A review and synthesis. Working paper #34, Center for Health Services and Policy Research, Northwestern University, Evanston, Illinois, 1980.

Other reports included in the literature reviewed, and papers which have been abstracted here, are cited in the text of this report.

3.3 Site Visits: Federal Agencies

The choice of federal offices to interview personally or by telephone was determined primarily by whether the office had responsibility to initiate, fund, execute, or review, evaluations and by the level of evaluation effort. User groups were determined by the target program for evaluations. The choice of individuals to interview within the office was determined primarily by determining whether they were knowledgeable about evaluation.

No central archive of federal evaluation experts exists. However, organizational charts and normal administrative data on contracts were very helpful. For instance, the USOE's Office of Evaluation and Dissemination has regularly produced a useful report on evaluation contracts, amounts, contractor, and contractor monitor. The project monitor list was a basis for many of the interviews within the Office.

The agencies in which staff members were interviewed included:

- U.S. Office of Education
    - Office of Evaluation and Dissemination
    - OED Division of Elementary and Secondary Programs
    - OED Division of Occupational, Handicapped, Developmental Programs
    - OED Office of the Assistant Commissioner
    - OED Division of Educational Replication
- National Institute of Education
    - Program on Dissemination and Improvement of Practice
    - Program on Testing, Assessment, and Evaluation
    - Office of the Director
- DHEW
    - Office of the Assistant Secretary for Planning and Evaluation
    - Office of the Assistant Secretary for Education
- Congressional Research Service
    - Division of Science Policy
    - Division of Education and Public Welfare
- Bureau of Education for the Handicapped
    - Division of Assistance to States
- U.S. General Accounting Office
    - Program Analysis Division
    - Human Resources Division

3.4 Site Visits and Telephone Surveys

Local and State Education Agencies

This section describes the procedures used to execute site visits to local and state education agencies, telephone surveys of the agencies, and the rationale for procedures.

Introduction

One field study component of the Project involved site visits and interviews at 6 state offices of education and 12 school district offices of evaluation. These were used as a vehicle for case studies. Telephone surveys of a sample of over 40 state offices of education, and of 200 school districts were also undertaken.

The main purposes of site visits and telephone surveys were to verify conclusions and implications of earlier research on the topic, to update what was known about local and state views of evaluation, and to assure that local and state views were represented in our report. Site visits are the basis for qualitative description, i.e., case studies. Telephone surveys are the basis for quantitative description.

Site Visit Protocol

The procedure for site visits to LEAs and SEAs involved:

(a) Initial contact with respondents by telephone,

(b) Follow-up letter to respondents outlining project and questions, confirmation of site visit date and visitors,

(c) Site visit and follow-up contact to clarify ambiguities.

The telephone contact and letter introduced the study in the following way:

> "The Project's main purpose is to review educational program evaluations supported by the federal government and to make recommendations for their improvement. The effort is mandated by Congress under Public Law 95-561 and is designed specifically to furnish appraisal independent of federal agencies and to examine evaluation at federal, state, and local levels of government. The Project is prospective in orientation, designed to provide: recommendations about evaluation policy and practice, the evidence to sustain recommendations, and the identification of issues and options. Questions to be addressed by the Project include:
>
> . Why and how are evaluations carried out?
> . What are the capabilities of those who carry out evaluations?
> . How are the results of evaluation used?
> . What recommendations can be made to improve policy or practice?

The main vehicles for providing answers to such questions are a critical examination of contemporary field research on each topic and field work by Project staff.

Current research which the Project will reanalyze includes major studies by commercial contractors, school districts, state offices of research, and federal agencies such as the U.S. Office of Education, National Institute of Education, U.S. General Accounting Office.

The product, a report to Congress, is scheduled for completion in June 1980. It will cover premises, recommendations, issues, options, and evidence bearing on evaluation policy. Information generated by the Project will be made available for competing analyses. Robert F. Boruch and David S. Cordray, of Northwestern University, have primary responsibility for direction of the Project and the final report.

The questions we'd like to discuss with your staff are identical to those outlined above: Why are evaluations done? What are capavilities of evaluators? How are results used: What recommendations can be made?

The responses to questions will not be attributed to individual respondents, unless the respondent prefers that he or she be identified. That is, individually identifiable responses will be maintained as confidential to the extent permitted by law."

Telephone Interview Protocol

The procedure for telephone interviews involved:

(a) Initial investigation to identify the evaluation officer within school district/state.

(b) Follow-up letter to respondents outlining project and questions.

(c) Telephone interview.

The information to be provided to the respondent in the initial contact and in the formal letter is identical to that given earlier for site visit interviews.

Selection of Local Education Agencies for Site Visits

The sampling frame used to select LEAs for site visits was defined as the largest 250 local school districts as determined through the National Center for Education Statistics: Education Directory: Public School Systems 1977-78. The ranks of each school size appear on pages 243-248 of this document. In total, 722 school districts are ranked, each of these having over 10,000 students enrolled during the 1976-77 school year.

The sampling frame was restricted to the larges 250 schools for the following reasons:

- We wanted sites that had a large enough enrollment to ensure that an evaluation office would be in place.

- We wanted sites that have multiple programs that would be in place.

Such a restriction regarding the sampling frame eliminates from consideration small school districts. The general feeling among the Northwestern staff was that we needed to examine evaluation practices within organizations (e.g. Offices of Evaluation and Research). Small districts would probably not have such an organizational structure. Some smaller districts have been included in the telephone phase of the project.

Sample size for site visits was restricted since the main purpose is to develop case studies and not statistical generalizations. The actual sample to be site visited was constructed in the following way:

1. 250 of the largest school districts were individually listed on 3x5 cards, organized by State within Region.

2. Sample size was set at 16.

3. Using a random start, every 15th card was selected. This process was repeated twice, generating two lists of 16 sites.

4. The two lists were treated as containing 16 pairs of sites and each pair was again randomly assigned to one of the two lists.

5. The two resulting lists were then randomly allocated to primary vs. secondary status. List 1 will be the basis for initial contact. In the event that we are not permitted to visit one of the sites listed in the primary list, its paired sites from list 2 will be contacted.

6. Any information about the particular LEA or SEA that we obtained through interviews has been included on the attached enumeration of sites.

7. Comparison of the sample with the distribution of Region and size of enrollment for the "population" reveals a closer correspondence.

8. The information that OED is compiling for us on amount and type of federal funds will be used as background information in preparation for the site visits.

The actual sites are listed in Table 1.

Table 1. Site Visits

LEA	DHEW Region	Enrollment	SEA
San Diego Unified School District, CA	9	125,463	California
Broward County, Florida	4	136,000	
St. Louis, Missouri	7	73,000	
St. Paul, Minnesota	5	38,105	Minnesota
Colorado Springs, Colorado	8	32,452	
Amarillo, Texas	6	25,899	Texas
Jersey City, New Jersey	2	34,698	New Jersey
Springfield, Massachusetts	1	27,892	Massachusetts
*Lansing, Michigan	5	28,984	Michigan
*Gaston County, North Carolina	4	34,759	
*Jefferson Parish, La	6	68,851	
Charleston, W. V.	3	45,428	

*Selected from alternate list. Three local education agencies were unwilling or unable to receive a site visit. They included: Norfolk, (Virginia), Wake County (North Carolina). and Shreveport, Louisiana. Shreveport was at the time dealing with court ordered desegregation. Pertinent staff members at Boise, Idaho, a fourth site, could not be reached during the inquiry period on account of a break between semesters.

Sample Selection for Telephone Interviews: School Districts

The design for this study involves a stratified random sample of 200 local school districts. The respondent in each case will be the person with major responsibility for program evaluation in the school district, often a director of research or of evaluation.

The main stratification variables for school districts are (a) size of student body and (b) level of federal funding for education programs. Population listings containing such information are available for 1976. We do not have the resources to update this listing and must assume that major changes are unlikely; the changes in the nature of federal funding of major programs over the past 4 years have not been substantial and this makes the assumption tenable. The sample size was selected to be of minimum size on the basis of confidence intervals for parameters in the aggregate sample. In particular, we determine sample size for proportion sampling based on the assumption that p = .10 for a .95% confidence interval (acceptable error margin of 5%). This is a simple approach, but we believe it will be sufficient to inform us about gross characteristics of the target group within our time and budget constraints.

The sample itself will be selected by taking every Nth school district in a stratified list, beginning with a random start, N being chosen to yield a sample size of 200. Replacement will be necessary for those districts undergoing transition, and will be taken randomly from the relevant stratum in the list.

Construction of Interview Questions

The construction of interview questions was guided by the general questions that we required to address by our contract. They were elaborated in discussion, put into a form suitable for interview work, pilot tested, and then revised. The questions outlined in the original work statement for the Project for the study were:

(a) Why and how are evaluations carried out (method and measurements)?

  (i) for whose information needs are evaluations carried out?

  (ii) what procedures are used to carry them out? How sound are they?

  (iii) what measurement techniques and devices are used and how appropriate are they to the program's objectives?

  (iv) are there conflicting requirements for different programs that lead to duplicative or unusually burdensome efforts?

(b) What are the capabilities including integrity and independence of those who carry out evaluations?

  (i) where are the evaluation activities located organizationally?

(ii) what is the background, training and experience of the staff and of those who carry out evaluations? If such services are provided by outside agencies, how are their services obtained and what is the nature of their relationship with in-house staff?

(c) How are the results of evaluation used?

(i) what are the conditions that facilitate or detract from their use?

(ii) when results are used, what is the nature of the changes they lead to? Can exemplary use be identified? Is their nature such that they can be adopted in other settings?

(d) What legislative, regulatory, funding or other changes might be proposed to improve the nature, conduct, and utility of evaluations?

Content of Questions

The specific questions addressed in the site visits are given in Tables 1, 2, and 3 of this Appendix.

A nearly identical set of questions about use of data were developed for each type of interviewee-Directors of Research and Evaluation, program administrators etc. at the school district and state levels. Questions were modified slightly to suit the context.

To make questions about use meaningful, they were constructed to refer to specific evaluation products generated for

- Title I
- Bilingual Education
- Vocational Education
- Special Education

Questions were further blocked into categories to recognize:

- Information required by federal reports
- Data collected in addition to that required by federal reports
- Formal evaluations initiated locally and bearing on each type of program
- Exemplary examples of use
- Federal and State reports on evaluation.

Pilot Tests

Interview questions and protocol for the site visits to local and state education agencies were developed at Northwestern University. They were field tested in January, 1980 through visits to the Office of Research of the School District of Columbus, Ohio and the Department of Education for the State of California.

Respondent Burden

It is partly in the interest of reducing the individual respondent's burden to the lowest possible level that oral responses, rather than written responses such as one might make on a questionnaire, are being elicited.

It is also in the interest of reducing the individual respondent's burden to a minimum that (a) only the questions implied by the Congress were put to respondents, and (b) information obtained from other surveys was used to reduce the need for probe questions. The latter includes, for example, surveys of research units in school districts by the UCLA Center for the Study of Evaluations.

It is in the interest of reducing respondent burden in the aggregate that sample size for the telephone survey has been reduced to a minimum subject to the constraint that a reasonable level of statistical precision is met.

The same applies to the number of case studies planned. We planned, at most, fifteen school districts and ten state offices of education. No formal statistical standards exist for judging adequacy of intensive case studies. But we believe that these provide informative case studies. Every effort was made to select sites so as to assure that school districts which have participated in related surveys during the past year did not fall into our sample.

3.5 Site Visits and Telephone Interviews: Title I Models

In order to provide a partial answer to the question "How are evaluations carried out?," a special effort was made to examine the use of Title I models at the local and state levels. The main objective was to catalog where models are being used, which models are used, and to describe the character of use. To assure that we understand the character of use, data generated by state and local officials on the basis of one model were also obtained. The main vehicles for this study were site visits and telephone interviews.

The ten Technical Assistance Centers in the Title I program were called to obtain information about the use of Title I models. State Title I coordinators and their evaluation personnel in ten states were also called on the same topic. At least twelve local education agencies were also contacted for information by telephone. The TAC coverage was complete. States were called only when TACs could not provide sufficient information or when we chose to pursue information in more detail. Only those local education agencies which use model C were called, based on referrals from the Technical Assistance Centers.

One site visit was made to the Region I (New England) Technical Assistance Center. Within Region I, the Providence Board of Education was visited. A second site visit was made to state education offices and three local education agencies in Florida on the basis of Florida's diversity in use of Title I models.

Complete data on the use of Title I models and testing data more generally were obtained from school districts in Providence, Rhode Island; Marion County, Florida; and Osceola County, Florida.

3.6 Roundtable Discussions

Informal roundtable discussions were developed to supplement the information generated from this study's case studies, surveys, and literature reviews. The participants, individuals who are generally knowledgeable about a particular topic, were invited to meet for discussions at Northwestern. The agenda for each was structured to invite opinion and evidence on questions this study was asked to address. The discussion topics, participants, and staff organizers are described below.

Roundtable on Community Reactions to Evaluation

On March 3, 1980, six respected and well-informed Evanston school parents discussed evaluation. They were selected based on their long-standing involvement with Evanston School District and their influence in the community. The school-community leaders who were invited include:

Barbara Zimmer - President, District 202 School Board
Alice Kreiman - President, District 65 School Board
Sharon Peterson - Member, District 65 School Board
Jessica Feldman - Head, Community-School Volunteer Force and Past President, Evanston PTA Council
Ethel Hilkevitch - Parent Advocate, Consultant for Special Education Evaluations
Mavis Hagemann - Head, School Education, Evanston PTA Council and Evaluation Consultant, Chicago Board of Education

District 65 is the Evanston K-8 elementary district. Current enrollment is 7,071 and the budget is about 20 million dollars. District 202 is the Evanston 9-12 high school district. Current enrollment is 3,821 with a budget of about 19 million dollars. Stephany Creamer organized the roundtable discussion.

Roundtable on Evanston School District Evaluation

Ida Lawler, Director of Testing and Evaluation at the Evanston School District 65, led a roundtable discussion with staff of this study on February 18, 1980.

Roundtable on Technical Assistance and Program Evaluation

Laura Crane, of Educational Testing Service, organized a roundtable discussion at ETS. Participants in the February 19, 1980 meeting included most staff of the Title I supported Technical Assistance Center at ETS, and its director, Ted Storlie.

Roundtable on Vocational
Education Program Evaluation

John Grasso, of the University of West Virginia, and former Congressman, Roman Pucinski were principal speakers at a roundtable discussion on evaluating vocational education programs on March 10, 1980. Janet Weeks organized the discussion.

Roundtable on
Utilization of Evaluations

The Roundtable took place March 14, at Northwestern. Its main purpose was to better understand current research on utilization of evaluation findings.

Participants included:

Dr. Jane David - President, Bay Area Research Associates. Dr. David is the author of a report for ASPE concerning school district uses of evaluation, and a history of Title I programs.

Dr. James DeGracie - Director of Research Office of Research and Evaluation, Mesa Public School District, Mesa, Arizona.

Dr. Frieda M. Holley - Director of Research and Evaluation, Austin Public School District, Austin, Texas, has studied utilization of evaluation in school districts.

Dr. Mary Kennedy - Project Director, the Huron Institute. Dr. Kennedy is principal investigator for a study of utilization of evaluations. Dr. Kennedy was formerly with the Bureau of the Educationally Handicapped.

Dr. Lee Sproull - Assistant Professor of Social Science, Department of Social Science, Carnegie-Mellon University. Dr. Sproull is principal investigator in a study funded by the Carnegie Corporation of school administrators' use of test scores.

Martin Bulmer - Professor, London School of Economics.

Hillel Weinberg - Congressional Staff, Office of Congressman Benjamin Gilman.

Theodore Storlie - Director, Technical Assistance Center, Region #5, Educational Testing Service.

Laura Leviton organized the round-table discussion with the assistance of Stephany Creamer.

Roundtable on Manpower

A roundtable focusing on issues related to the capabilities of educational evaluators and the factors involved in managing and conducting evaluation research was held May 9th at Northwestern. The purpose of the roundtable was to elicit experts' reactions to findings and to proposed recommendations on evaluation manpower, training, and management of research at the federal, state, and local levels. The individual participants were:

Dr. Launor Carter - Vice President, Systems Development Corporation, Santa Monica, California. A major contractor for federal evaluations, Dr. Carter and his staff have been responsible for conducting such projects as the Longitudinal Evaluation of the Emergency School Assistance Act Pilot Program, the Evaluation of the Title I Program in State Institutions for Neglected and Delinquent Children, and the Evaluation of the Sustaining Effects of Compensatory Education.

Dr. Harrison Fox - Visiting Professor, Industrial College, Washington, D.C. Dr. Fox is co-author of the book Congressional Staffs: The Invisible Force in American Lawmaking which in part focuses on the capabilities of congressional staffers and their associated activities.

Dr. Alexander Law - Deputy Superintendent for Research and Evaluation, State Department of Education, Sacramento, California. Dr. Law's expertise is in the area of state educational evaluation and the management of these activities, and he is the author of several papers on these topics.

Dr. Len Nachman - Evaluation Supervisor, Office of Planning and Evaluation, State Department of Education, St. Paul, Minnesota. Dr. Nachman is currently investigating the practiceof outside contracting at the state level and the development of appropriate guidelines for the selection of qualified contractors.

Ms. Laure Sharp - Bureau of Social Science Research, Washington, D.C. Ms. Sharp is coauthor of an early study on the federal contracting process for social program evaluation and the more recent project on identifying the characteristics of those organizations currently responsible for educational research, development, dissemination and evaluation.

Dr. Carol Tittle - School of Education, University of North Carolina, Greensboro, North Carolina. Dr. Tittle has conducted vocational education evaluations for the state of New York and is one of the co-authors of the Handbook of Vocational Education Evaluation published by Sage Press. She is currently a professor in the master's program in educational evaluation at Greensboro.

Georgine Pion organized the roundtable with the assistance of Stephany Creamer and Lucina Gallagher.

3.7 Formal Presentations and Discussions

In the time available, the Project staff was able to make only a few formal presentations of some results of the study. Our main interest in doing so was to elicit professional criticism of the character of results and recommendations. The meetings at which we discussed parts of the findings include:

1. American Educational Research Association Annual Meeting, Task Force on Meta-analysis, Boston, April 8, 1980.
2. Special National Workshop, Research Methodology and Criminal Justice Program Evaluation, Baltimore, March 17, 1980.
3. National Academy of Sciences, Committee on Program Evaluation, Washington, D.C., March, 1980 and Evanston, Illinois, June, 1980.
4. International Association for Social Science Information Service and Technology, Annual Conference, Washington, D.C. May 2-4, 1980.

3.8 Machine Based Searches

National Institute of Education Machine Based Search

A search of all grants and contracts bearing on program evaluation was initiated using the NIE's computer based "Program Management System." The search was based on key words such as: evaluation, student evaluation, effects, effectiveness, program evaluation, achievement, gains, data bases, analysis. Information on grant obligations for fiscal year 1977 through 1979 was available. We were unaware of the existence of the system, despite interviews with NIE staff, until March, 1980. We should have informed ourselves about it earlier. We are grateful to Dr. Lois-ellen Datta of NIE for her assistance in doing so.

We attempted to classify grants and contracts using the categories we have used routinely for classifying types of evaluation research: needs, assessment, process/implementation, outcome. Much of the research covered two of these areas simultaneously and a good deal of it had to be classified into a fourth broad category: technical research, support, and development. So for example a study at New York's Empire State University of a management information system was classified as bearing most directly on process/implementation. Research on development of standards was classified as technical research, support, and development. A training program for women at San Francisco in education was classified as technical research, support, and development. A study of determinants of career entry at Johns Hopkins was classified as exploratory/needs and as outcome evaluation research.

Legal Search

The LEXIS computerized legal document retrieval system was used to search the United States Code for statutes bearing on evaluation. The search employed joint occurrence of the key words education and evaluation as the main criterion for enumerating statutes. Scope of the search was narrowed further by confining attention to Title 20 of the Code, containing statutes relevant to public education. Joe S. Cecil, an attorney with a Ph.D. in methodology and evaluation research carried out the investigation. His report, describing the search strategy, difficulties encountered in using LEXIS, and results is given in an Appendix.

Educational Resources Information Center (ERIC)

ERIC and other similar computerized bibliographic services (NTIS, Smithsonian) were used by several project staff primarily to identify important journal articles, government reports, and research activities bearing on the Project's objectives. Searches typically focused on subjects such as evaluation of vocational education programs or the use of evaluations in public policy making. A considerable amount of useful information, particularly in the form of government reports and documents, was generated in these searches. Unfortunately, there is no good way for users to estimate the coverage of these subject-area searches. This is true not only in the sense of documents not found by the particular search strategy employed but also in terms of the selection process for including documents in the databases.

A second major use of ERIC was for identifying and retrieving specific selected documents. ERIC is the best database in this regard because all documents listed are available on microfiche, many are also available in copies. Many major libraries, including the Northwestern University library maintain microfiche collections of all ERIC documents. Copies can also be obtained either by mail or by computer order directly from ERIC. In most cases, documents which are not available in paper copy are of such poor quality that reproduction is not possible. This is a problem which could easily be remedied.

This type of selective search can be done in a number of different ways including by author, title, sponsoring or funding agency or federal contract number. For example, a sample of 50 recent evaluation contracts of USOE's Office of Evaluation and Dissemination were selected for review. ERIC was searched, using the contract number, to identify and retrieve the reports associated with each contract. Slightly less than half of the contracts (42%) showed no ERIC citations. However at least a third of these contracts did have reports available through ERIC based on a listing obtained from USOE. In total, 35 of the 50 evaluation contracts did have reports catalogued in ERIC. This example illustrates both of the coverage problems described earlier. It also suggests that steps need to be taken to standardize entries within the system and to examine the process by which government reports and documents are acquired and selected for inclusion in the ERIC system. The ERIC searches were executed by Roger Straw.

3.9 Evasive and Otherwise Diverting Responses

1. In response to the question: Can we have the report?

   - It's upstairs
   - We only printed a few and I don't have any left.
   - I'll send it to you first thing in the morning.
   - Gee, we don't issue them in hard cover.

2. In response to the question: How do you use evaluation results?

   - Let me tell you about our program.
   - I use them all the time.
   - Let me tell you about the history of this city.

3. In response to the question: Can you provide a citation to the studies you referred to earlier?

   - Let me tell you about the testimony I've heard.
   - It's in the Hearings, sometime during 1978-79, in one of the 30 odd volumes.

4. In response to the question: Who asks for these evaluations?

   - The bureaucrats (if responent is an ex-Congressman).
   - The Congress (if respondent is a bureaucrat).
   - The "feds" (If respondent is a local school district person).

5. Program director's reaction to an evaluator's report that the program has little or no detectable impact:

   - The evaluator is inexperienced, or doesn't understand the program.
   - The methodology is wrong: tests are irrelevant.
   - It came too late to use in our decision
   - Let's see if we can find another evaluator.

6. Can you provide evidence that the report was useful?

   - Everybody said it was useful.
   - Ask me another question.
   - Let me answer another question.

APPENDIX 3

7. In answer to the question: How do you know your program is effective?

   - Let me tell you about little Mary.
   - Let me tell you about little Sue.
   - Let me tell you about Jack who participated in the program and succeeded and Jill, his sister, who did not participate and is in jail.

8. In response to the question: What do you think of bureaucrat Y's efforts to evaluate?

   - Y is a fine and decent man, a good friend, and neighbor. But he's been wrong for the past 10 years (If a bureaucrat responds).
   - He's been wrong for the past 10 years (if Congressional assistant).
   - He's been right for the past 10 years (if Congressional assistant).
   - Who is he? (If parent, principal, etc.)

9. How do you use evaluations?

   - We use evaluations like a drunk uses a lamp post. . . for support rather than illumination.
   - You mean we're supposed to use them?

Appendix 4.

Tabular Comparison
of Standards for Evaluation

Roger B. Straw and David S. Cordray

The attached table is an attempt to make a comparison between six sets of standards or criteria which have been proposed for judging the adequacy of evaluations. The standards proposed by the Joint Committee on Standards for Educational Evaluation are listed in column 1 and are used as the basis against which to compare the other standards. The main reasons for choosing the Joint Committee's classification are that they also provide detailed descriptions and analysis of each standard along with examples of good and bad practice and that each standard can be reasonably described with the indicated descriptor. The descriptors are presented with their extended definitions immediately following the table. Each of the other five columns compares another set of standards to the Joint Committee's. If there is an entry, it indicates that the standard is roughly equivalent to the corresponding Joint Committee standard. If there is no entry, it means that no explicit recognition of that aspect was apparent. In columns 3 through 5, the number refers to the standard number listed in the original document. The numbers along with their associated content are reproduced following the table. The standards used by the Joint Dissemination Review Panel in judging the adequacy of evaluation results presented to them as support for claims of program effectiveness do not lend themselves to numerical expression. Because of this, the asterisks in column 6 are meant to indicate that the JDRP considers these particular aspects of evaluations important in making their decisions.

Joint Committee[1]	Phi Delta Kappa[2]	CSE[3]	ERS[4]	GAO[5]	JDRP[6]
Accuracy Standards	Scientific Criteria				
Described Object	External validity	1,2	1,6	A2,B1	*
Described Context	External Validity				*
Described Purposes and Procedures	Internal Validity	3	3,10,13,14, 20,21,49,50	A1,A3,A6,B1, D2,E1,E2	*
Described Info. Sources	Objectivity	5,6	15,17,31	A4,E1,A6	*
Valid Measurement	Internal Validity		16,17,23	A3,A5,B1	*
Reliable Measurement	Reliability		18,17,23	A5	*
Systematic Data Control			19,20,21,22,24, 25,26,30,50	B1,B2	
Analysis of Quant. Information		9	26,32,33,34, 35,36,37,38	B2,C1	*
Analysis of Qual. Information			32,33,34,35, 36,37		*
Justified Conclusions		11,12	39,46	A3,D1	*
Objective Reporting	Objectivity	10	40,41	D1,D2	
Utility Standards	Practical Criteria				
Evaluator Credibility	Credibility		6,12	A7,A8	*
Audience Identification	Relevance	4	2,18	A1,A8	
Information Scope and Selection	Scope Importance	7		A1,A7,A8	*
Report Clarity		13	41,42,46	D1,D2,D3	*

Joint Committee	Phi Delta Kappa	CSE	ERS	GAO	JDRP
Report Dissemination	Pervasiveness		47,48,54,55	D2	
Report Timeliness	Timeliness		5,6,51	A7	
Evaluation Impact	Pervasiveness		5,6,52,54,56	D3	*
Propriety Standards					
Formal Obligation			5,7,9,10,11	A7,A8,E2	
Conflict of Interest			6,8	A7	
Full and Frank Disclosure			41,44	D1	
Public's Right to Know			6,9,30,48,49	A7,E2	
Rights of Human Subjects			6,9,28,29	A7	
Human Interactions			29		
Balanced Reporting			43	D1	
Fiscal Responsibility			4		
Feasibility Standards	Prudential Criterion				
Practical Procedures	Efficiency		27		
Political Viability			6	A7	
Cost Effectiveness	Efficiency		6,10	A7	

---

[1] Joint Committee on Standards for Educational Evaluation, 1978.
[2] Phi Delta Kappan National Study Commission on Evaluation, 1971.
[3] Center for the Study of Evaluation, Standards for reviewing written reports, 1978.
[4] Evaluation Research Society Draft Standards, 1979.
[5] U.S. Government Accounting Office Checklist for assessing Impact Evaluations, 1978.
[6] Joint Dissemination Review Panel Ideabook, 1977.

Joint Committee on Standards
for Educational Evaluation: Draft Standards, 1978

Descriptor	Definition
Described Object:	The object of the evaluation (program, project, material) should be described so that it is clear what form(s) of the object is (are) being considered in the evaluation.
Described Context:	The context in which the program, project, or materials exist(s) should be described in enough detail so that the likely influences of the context on the object may be identified.
Described Purposes and Procedures:	The purposes and procedures of the evaluation should be described in enough detail so that they can be identified and assessed.
Described Information Sources:	The sources of information should be described in enough detail so that adequacy of the information can be assessed.
Valid Measurement:	The information-gathering instruments and procedures should be chosen or developed and implemented in ways that will assure that the interpretation arrived at is valid for the given use.
Reliable Measurement:	The information-gathering instruments and procedures should be chosen or developed and implemented in ways that will assure that the information obtained is reliable.
Systematic Data Control:	The data used in an evaluation should be reviewed and corrected so that evaluation reports will not be needlessly flawed.
Analysis of Quantitative Information:	An evaluation's quantitative information should be appropriately and systematically analyzed so that supportable interpretations are enabled.
Analysis of Qualitative Information:	An evaluation's qualitative information should be appropriately and systematically analyzed so that supportable interpretations are enabled.
Justified Conclusions:	The conclusions reached in an evaluation should be explicitly justified so that the audiences can assess them.

Descriptor	Definition
Objective Reporting:	Evaluators should be independent of what is evaluated, and the evaluation procedures should provide safeguards so that the evaluation findings and reports are not distorted by the personal feelings and biases of any party to the evaluation.
Evaluator Credibility:	The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation so that their findings achieve maximum credibility and acceptance.
Audience Identification:	Audiences involved in or affected by the evaluation should be identified so that their needs can be served.
Information Scope and Selection:	Information collected should be of such scope and selected in such ways that pertinent questions about the object of the evaluation are addressed and so that the information is responsive to the needs and interests of specified audiences.
Report Clarity:	The evaluation report should clearly describe the object being evaluated and its context and the purposes, procedures, and findings, so that the audiences will readily understand what was done, why it was done, what information was obtained, what conclusions were drawn, and what recommendations were made.
Report Dissemination:	Evaluation findings should be disseminated to clients and other right-to-know audiences so that they can assess and use the findings.
Report Timeliness:	Release of reports should be timely so that audiences can best use the reported information.
Evaluation Impact:	Evaluations should be planned and conducted in ways that encourage follow-through by members of the audiences.
Formal Obligation:	Obligations of the formal parties to an evaluation (what is done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.

Descriptor	Definition
Conflict of Interest:	Conflict of interest should be avoided in evaluation, or, should it occur, dealt with openly and honestly so that it does not compromise the evaluation processes and results.
Full and Frank Disclosure:	The evaluator's oral and written reports should be open, direct, and honest in their disclosure of pertinent findings, including the limitations of the evaluation.
Public's Right to Know:	The formal parties to an evaluation are responsible to uphold the principles of the public's right to know, within the limits of other related principles and statutes, such as those dealing with public safety and the right to privacy.
Rights of Human Subjects:	Evaluators must design and conduct their evaluations in such ways that the rights and welfare of the human subjects are respected and protected.
Human Interactions:	Evaluators should be respectful of human dignity and worth in their interactions with other persons associated with an evaluation.
Balanced Reporting:	The evaluation should be equitable in its presentation of strengths and weaknesses of the object under investigation so that strengths can be built upon and problem areas addressed.
Fiscal Responsibility:	The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and should be otherwise prudent and ethically responsible.
Practical Procedures:	The evaluation procedures should be practical so as to ensure that associated disruption is kept to a minimum, and that findings can be obtained.
Political Viability:	Evaluations should be planned and conducted with anticipation of the different positions of various interest groups so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.
Cost Effectiveness:	The evaluation should produce information of sufficient value to justify the resources used.

Center for the Study of Evaluation: Standards for Reviewing Evaluation Reports, 1978

Standard Number	Standard Content
1	The program or product or other object under study in the evaluation is described so that its objectives are clear.
2	The program or product or other object under study in the evaluation is described so that the form of its actual implementation is clear.
3	The purposes of the evaluation are described; purposes may be stated in terms of the evaluation questions or objectives.
4	Audience(s) for the evaluation information are identified.
5	Participants in the educational program and the evaluation study, and how they were selected for participation, are described.
6	Data collection sources, such as tests, records, or observation forms, are identified.
7	The data collection sources are comprehensive enough to answer the evaluation questions.
8	The reliability of the data collection sources, and the validity of the data collection sources for the purposes intended is described.
9	Data analysis procedures are described or are evident (as in detailed tables).
10	Evaluation results are described or presented.
11	Conclusions or recommendations are drawn from the results.
12	The congruence of the conclusions with the information provided is described or evident.
13	The written presentation of whatever was done in the evaluation is clear (even if standards above were not met).

Evaluation Research Society
Draft Standards

Standard Number	Standard Content
1	The purposes and characteristics of the program or activity to be addressed in the evaluation effort should be specified as precisely as possible.
2	The decision makers and potential users of the evaluation results should be identified and their expectations made clear.
3	The type of evaluation effort required should be identified and its objectives made clear; the range of activities to be undertaken should be specified.
4	An estimate of the cost of the proposed evaluation effort and, where appropriate, alternatives should be provided; this estimate should be prudent, ethically responsible, and based on sound accounting principles.
5	Agreement should be reached at the outset that the evaluation is likely to produce information of sufficient information value, applicability, and potential for utilization to justify the resources used.
6	The feasibility of undertaking the evaluation effort should be estimated either informally or through formal evaluability assessment.
7	Restrictions, if any, on access to the data and results from an evaluation should be clearly established and agreed to between the evaluator and the client at the outset.
8	Potential conflicts of interest should be identified and dealt with openly and honestly. Should the possibility of conflict of interest occur, steps should be taken to avoid compromising the evaluation processes and results.
9	Respect for and protection of the rights and welfare of all parties to the evaluation should be a central consideration in the negotiation process.
10	Accountability for the technical and financial management of the evaluation once it is undertaken should be clearly defined.

Standard Number	Standard Content
11	All agreements reached in the negotiation phase should be specified in writing, including the obligations of all formal parties to the evaluation and of all parties with specific roles to play in the effort.
12	Evaluators should not accept obligations that exceed their personal qualifications or the resources available to them.
13	For all types of evaluations, a clear methodological approach or design should be developed and justified in order that rival explanations and threats to the validity of conclusions and inferences can be anticipated.
14	For impact studies, the central evaluation design problem of estimating the effects of non-treatment, and the choice of a particular method for accomplishing this, should be fully described and justified.
15	If sampling is to be used, the details of the sampling methodology (choice of unit, method of selection, time frame, etc.) should be explained, justified, and based on an explicit analysis of the evaluation's requirements, including generalization.
16	The measurement methods and instruments should be specified and described, and their reliability and validity estimated for the population or phenomena to be measured.
17	Professionally outmoded or discredited procedures and instruments should not be specified for use.
18	The necessary cooperation of program staff, affected institutions and members of the community, as well as those directly involved in the evaluation, should be planned for and assurances obtained.
19	A comprehensive data collection and preparation plan should be developed in advance of any data collection.
20	Given a sound design, the data collection plan should conform to it. However, if the development of the data collection plan produces insights into realities that might compromise the design, appropriate revisions of the design should be made before proceeding.

Standard Number	Standard Content
21	Provision should also be made for the detection, reconciliation, and documentation of departures from the original design that may occur during data collection.
22	Evaluation staff should be selected, trained, and supervised according to criteria that ensure competence, consistency, impartiality, and ethical practice.
23	The validity and reliability of data collection instruments and procedures should be verified under the prevailing circumstances of their use.
24	There should be an analysis of the sources of error that need to be taken into account, and provisions for quality assurance and control should be established that are collectively adequate to meet the requirements of the overall design and anticipated analyses.
25	The data collection and coding procedures should provide safeguards so that the findings and reports are not distorted by the personal feelings and biases of data collectors.
26	Data from secondary sources should be verified.
27	Data collection activities should be conducted with a minimum of disruption and imposition in the program or other settings where the data collection activities take place.
28	Procedures that might entail significant adverse effects or risks should be subjected to external review; if approved, informed consent should be obtained in advance of their application.
29	All data collection activities should be conducted so that the rights, welfare, dignity, and worth of individuals are respected and protected.
30	Data should be handled and stored so that unintentional release is prevented and so that access to individually identifying data is as limited as possible.
31	Documentation should be provided of the source, method of collection, circumstances of collection, and processes of preparation for each item of data.

Standard Number	Standard Content
32	The analytic procedures should be matched to the general purposes of the evaluation effort, the design, and the data collection.
33	All analytic procedures, along with their underlying assumptions and limitations, should be described explicitly, and the reasons for choosing the procedures should be explained.
34	Analytic procedures should be appropriate to the properties of the measures used and to the quality and quantity of the available data.
35	The units of analysis should correspond to the units of assignment and comparison.
36	No outmoded or discredited analytic procedures should be employed.
37	The analyses should be replicable by other qualified evaluators.
38	When quantitative comparisons are made (e.g., X is greater than Y), some indication of the confidence that can be placed on the stated differences and some indication of their magnitude or consequence should be provided.
39	Cause-and-effect interpretations should be bolstered not only by reference to the design but also by recognition and elimination of plausible rival explanations.
40	Purported "findings" should be reported in a manner that distinguishes fact from speculation and objective inference from subjective interpretation.
41	Findings and recommendations should be presented clearly, completely, and fairly. (See standard 40.)
42	Findings and recommendations should be organized and stated in forms that are conducive to understanding by the intended audience and address their decision-making requirements.
43	Findings and recommendations should be presented in a framework that indicates their relative importance.

Standard Number	Standard Content
44	Assumptions and limitations should be explicitly acknowledged.
45	Those issues and unanswered questions that need further study should be identified.
46	Complete explanation and description of how findings and results were derived should be accessible.
47	Persons, groups, and organizations who have made contributions to the evaluation effort should receive feedback appropriate to their perspectives and possible applications.
48	Disclosure should follow the legal and proprietary understandings agreed upon in advance (standard 7), with the evaluator serving as a proponent for the fullest, most open disclosure possible.
49	Policies and procedures on access to the data should be formulated and made available; these should specify the officials authorized to release data and the criteria for release.
50	The finished data base and associated documentation should be organized in a manner consistent with the accessibility policies and procedures. (See standards 30, 31, and 33.)
51	Evaluation results should be made available to appropriate users before the relevant decisions and discussions about the program must be undertaken.
52	Users of evaluation results should be encouraged and helped to undertake further explorations of those issues and questions not resolved by the evaluation effort. (See standard 45.)
53	Evaluators should try to anticipate possible misinterpretations and misuse of evaluative information and should provide safeguards to diminish the improper use of evaluative information.
54	Within the limitations of the initial understandings about disclosure, activities to further utilization should be designed with consideration for any broader implications that the evaluation may have for secondary audiences.

Standard Number	Standard Content
55	The evaluator should bring to the attention of decision makers any important program effects associated with the evaluation process.
56	In the utilization process, evaluators should differentiate clearly between their activities as change agents and their roles as relatively impartial scientists.

U.S. GAO Exposure Draft: Assessing Social Program Impact Evaluations: A Checklist Approach, 1978

Standard Number	Standard Content
A1	Have evaluation goals been defined and described?
A2	Has the evaluability of the program been determined?
A3	Has a clear evaluation approach been developed and justified, and potential threats to the validity of conclusions and inferences anticipated and accommodated?
A4	Has a method for sample selection been explained and justified?
A5	Have measurement methods been identified and their validity and reliability assessed?
A6	Have the frequency and timing of measurements been specified and explained?
A7	Has the feasibility of performing the evaluation been examined?
A8	Has the necessary cooperation been obtained?
B1	Have procedures for quality control of data been identified and implemented?
B2	Have preliminary analyses been performed to detect missing or inconsistent information and correct deficiencies in the study plan?
C1	Have the statistical methods and model for use in the analysis and the rationale for their selection been specified?
C2	Has the unit of analysis been justified?
C3	Have the assumptions essential to statistical methods and model been specified and have their conditions been met?
D1	Have the findings been presented clearly, completely and fairly?
D2	Have specific procedures been used to assure the report's quality?

Standard Number	Standard Content
D3	Have follow-up provisions been made to assist decision makers in using the report?
E1	Has adequate documentation of the evaluation been maintained?
E2	Has a procedure been established for release of data for audit, reanalysis and other evaluations or research?

Brief Bibliography on Guidelines and Standards to Facilitate the Appraisal of Evaluative Research

Auditor General of Canada. Evaluation Standards: Efficiency issues and effectiveness evaluation. In *100th Annual Report to the House of Commons*. Hull, Quebec: Printing and Publishing Supply and Services, 1978

Bailar, B. A. and Lanphier, C. M. *Development of survey methods to assess survey practices*. American Statistical Association, Washington, D.C., 1978.

Baron, J. B. and Baron, R. M. In search of standards. In R. Perloff and E. Perloff (Eds.), *New Directions for Program Evaluation: Values, ethics, and standards*. San Francisco: Jossey-Bass, in press.

Davis, H. R., Windle, C., and Sharfstein, S. S. Developing guidelines for program evaluation capability in Community Mental Health Centers. *Evaluation*, 1977, *4*, 25-34.

Emrich, R. L. Proposed evaluation guidelines and standards. Sacramento, Calif.: California Council on Criminal Justice, 1973.

Gordon, G., and Morse, E. V. Evaluation research. In A. Inkeles (Ed.), *Annual Review of Sociology*, 1975 (Vol. 1).

Joint Committee on Standards for Educational Evaluation. Standards evaluations of educational programs, projects, and materials (Draft). Kalamazoo: Western Michigan University Research Center, 1979.

McTavish, D. G., Cleary, J. D., Brent, E. E., Perman, L., and Knudsen, K. R. Assessing research methodology: The structure of professional assessments of methodology. *Sociological Methods and Research*, 1977, *6*(1), 3-44.

Ridings, J. M. Standard setting in accounting and auditing: Considerations for educational evaluation. Ph.D. Dissertation. Kalamazoo: Western Michigan University, 1980.

Scriven, M. Checklist for the evaluation of products, producers, and proposals. Unpublished manuscript, San Francisco State University, Undated.

Tallmadge, G. K. *The joint dissemination review panel IDEABOOK*. Washington, D.C.: National Institute of Education, October, 1977.

U. S. General Accounting Office. *Evaluation and analysis to support decision-making*. Washington, D.C.: U. S. GAO (PAD-26-9), December, 1975.

U. S. General Accounting Office. Assessing social program impact evaluations: A checklist approach. (PAD-79-2). Washington, D.C.: U. S. General Accounting Office, October, 1978.

U. S. General Accounting Office. *Status and Issues: Federal program evaluation*. Washington, D.C.: USGAO, October 1978.